SIRPI
Visualize and Decide

INSOFE
Inspire...Educate...Transform.

ART & SCIENCE OF STORY TELLING WITH DATA VISUALIZATIONS

Dr. Anand Lakshmanan
CEO, SIRPI PRODUCTS AND SERVICES PRIVATE LIMITED
@lan24hd
+91 83107 64903
anand@sirpi.io

My Profile

- Ph.D in Electromagnetics & Antenna Design
- Ex - Apple Design Engineer
- 18 yrs of experience (academia + industry)
- 7 Patents
- Now into :
  - Data Science Training
  - Data Visualization Consulting

My Journey

- Cognizant, Chennai, India
- Clemson University, Clemson, USA
- University of Wisconsin, Madison, USA
- Southern Methodist University, Dallas, USA
- Apple, Cupertino, USA
- Sirpi, Bengaluru, India

My Learnings

- World needs problem solvers
- Old paradigm : Use theoretical knowledge to solve problems
- New paradigm : Collect data, visualize data, propose solutions

DATA SCIENCE

ART & SCIENCE OF DATA VISUALIZATION

ART

Helps you remember

Connects with your heart

Kindles emotion

Inspiration

Design

ART

Out of box

Beauty

Energizes

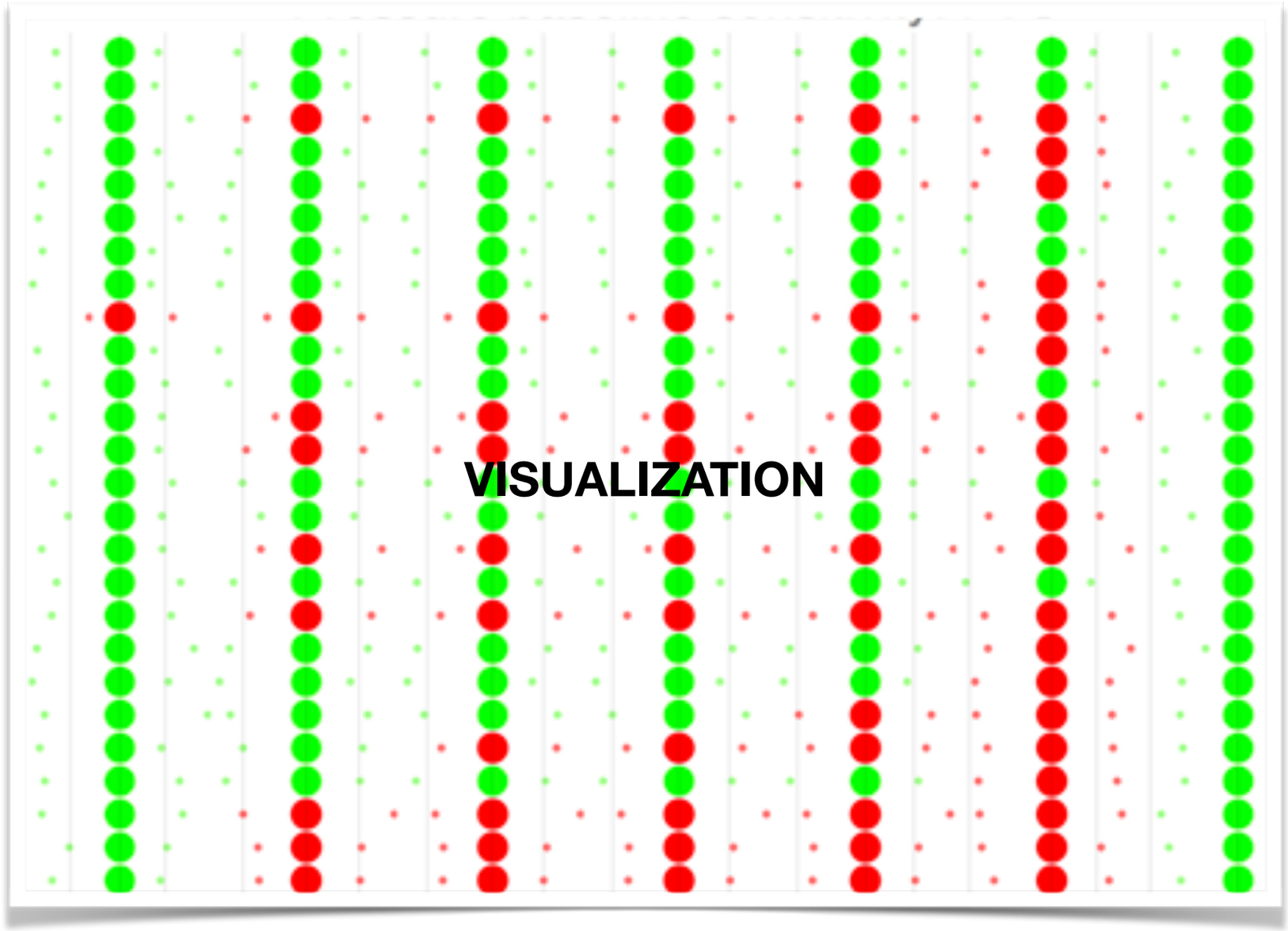
Off road

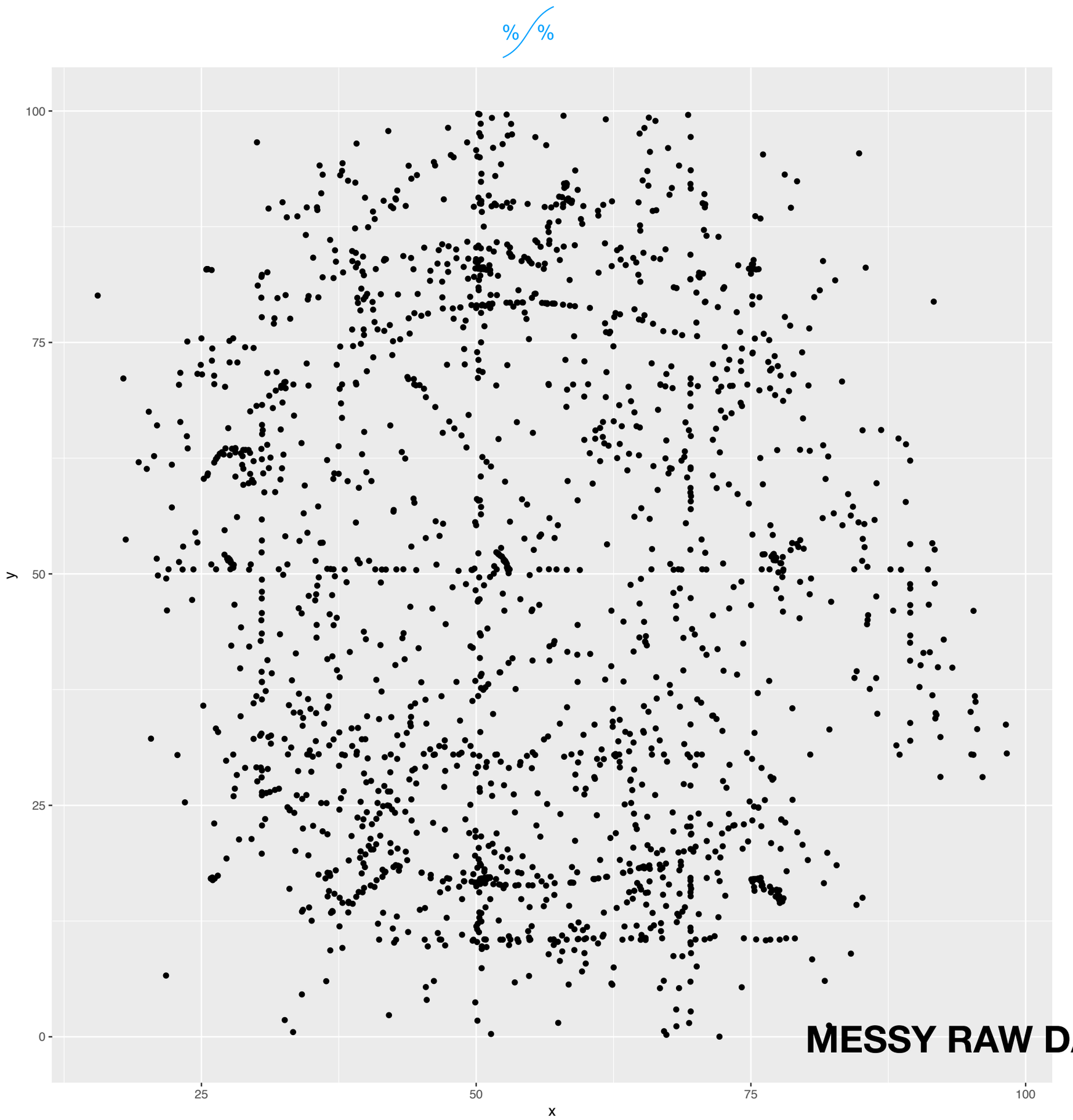
This ?

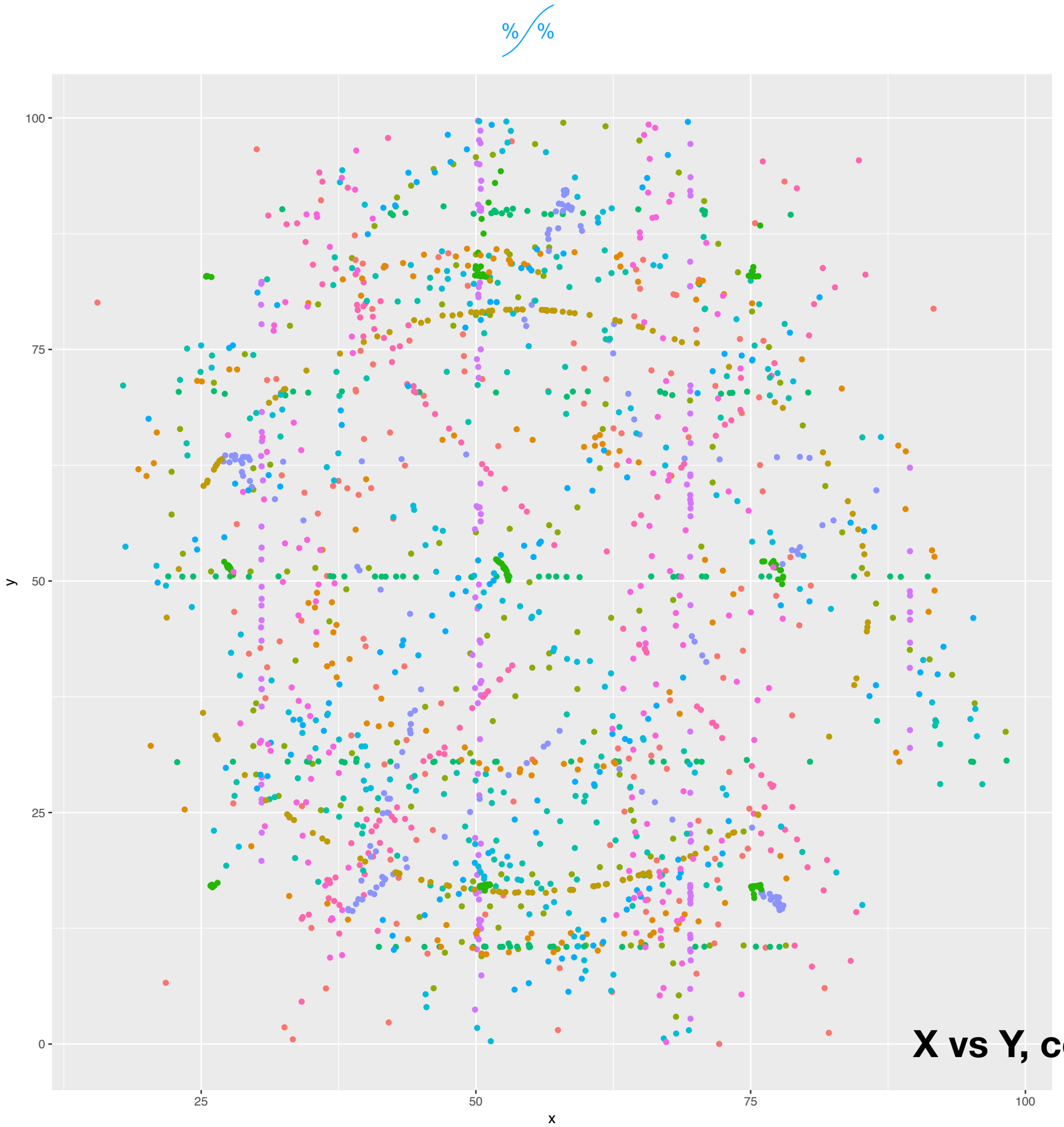
f01	4.1749191781406	0.15455894200499998	3.702082253813509	3.280027722979433	0.06741096828976234
f02	3.8395846340118	0.16476975431800023	4.2913432056800405	3.4786189677219115	0.11918208935102026
f03	1.6383606029726001	0.12677683505100013	7.738030005175871	1.3802801553772217	0.06128564571183692
f04	1.7909154464608332	0.07022682691699988	3.9212809882108366	1.5467582715516506	0.05859430013147149
f05	2.8946250018836	0.14400622755699954	4.974952799180942	2.202101132380544	0.04476651230889628
f06	1.8232893974498	0.10271569613099985	5.633537729921883	1.3005688459565456	0.06072515412097301
f07	1.8269532432624	0.04861481164499981	2.6609773306616256	1.2880870250891847	0.01802114871234939
f08	1.7846759221886	0.07439246663699972	4.168401988959994	1.2017291189882398	0.0431324494202856
f09	2.7434286624924	0.14897372429900013	5.430202225985996	2.282360210752601	0.07459418956779329
f10	2.3481929137021997	0.06024658855000009	2.565657540249294	2.1022338325252523	0.07620932107805611
f11	2.763768398908	0.2200342533949997	7.961385385328887	2.100227881564459	0.07059657795876229
f12	5.1264973542400005	0.16337622368199956	3.1868976494618706	4.243360128845416	0.10574556713767347
f13	3.0175604929284	0.10261981958400046	3.4007543452563156	2.3629857759170148	0.07298021584006875
f14	3.0461130895601665	0.14378295911300043	4.[illegible]605543	2.370270702073574	0.040570802030806874
f15	3.5474003959963336	0.21505123510700042	6.062220530552782	3.0519798239632143	0.08498396306072609
f16	2.9060571990048	0.07706928671399949	2.6520223600689077	2.306750350700985	0.04256490357557219
f17	2.9414247275436	0.04083499071499963	1.3882725038862767	2.754675361006785	0.028036122177582
f18	4.5720515996628	0.16308494399600004	3.5669970130701922	3.99762811177299	0.14382704039224903
f19	3.6043237033798	0.08456529803999979	2.346218181255538	2.9669194236092267	0.018778781117449128
f20	5.331176432757	0.06786739876700043	1.273028563639246	4.3972488918881165	0.03339472081987793
f21	1.4286951577528	0.027021461506000044	1.8913384957853572	1.2265395508091264	0.04688161211614039
f22	2.0609493811754	0.09454002821899987	4.587207676351655	1.6971471060275023	0.06034120474063309
f23	4.272702461729001	0.09757922061500057	2.283782254650935	3.348270191763526	0.04886480399686155
f24	2.8776615701505	0.10902804545699984	3.7887723347293316	1.9573265997582214	0.08051001243593348
f25	5.833884322186	0.1368772414889996	2.34624538180268	5.393706337002649	0.09804550343722696
f26	4.8300593608878	0.09359876356499974	1.9378387835754387	4.048524432599245	0.06223400797501277

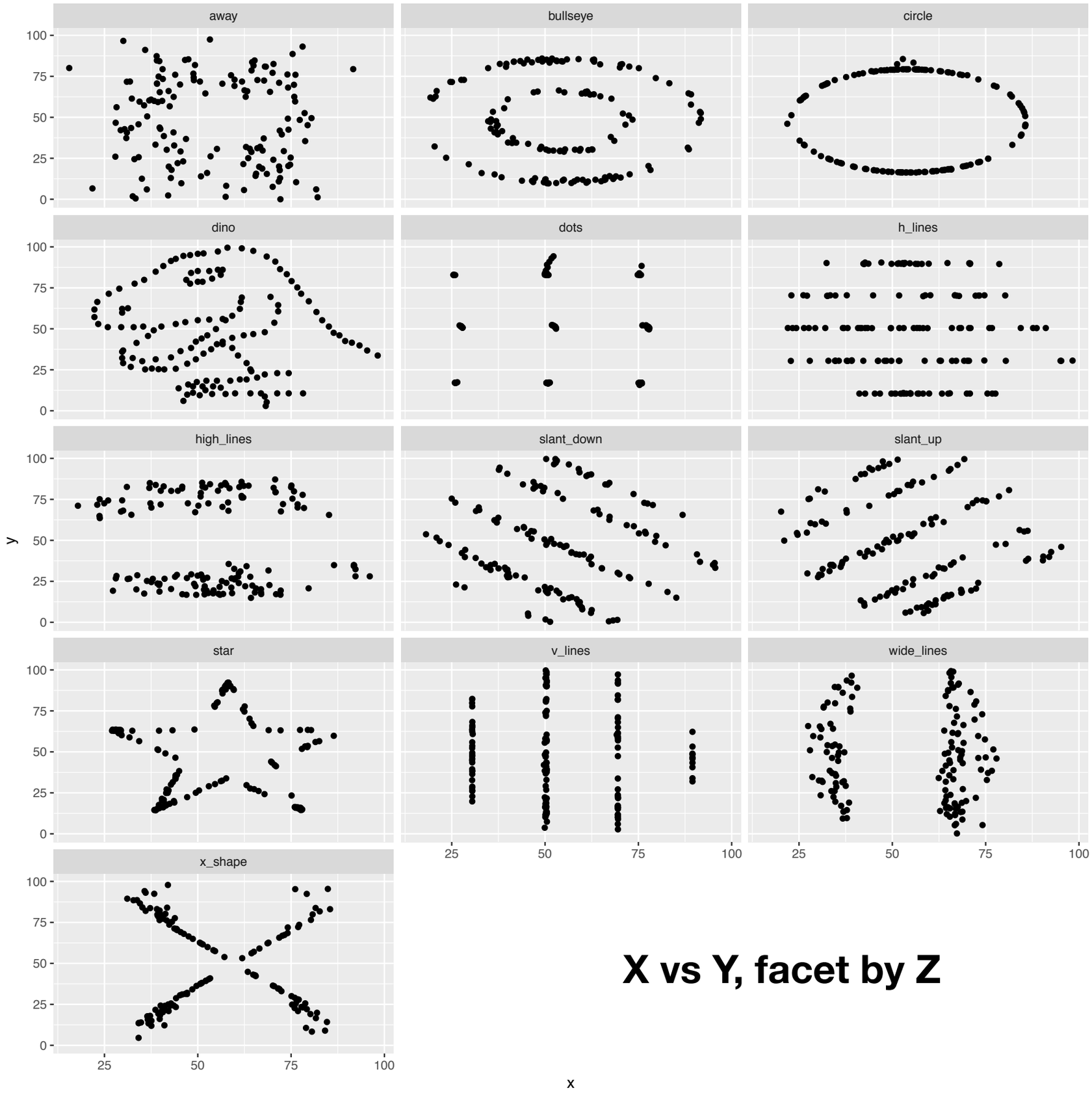
DATA

Or this ?




%/%
100
75
50
25
0
y
25
50
75
100
x
MESSY RAW DATA X vs Y


%/%
100
75
50
25
0
y
25
50
75
100
x
X vs Y, colour by Z


away
bullseye
circle
dino
dots
h_lines
high_lines
slant_down
slant_up
star
v_lines
wide_lines
x_shape
100
75
50
25
0
y
25
50
75
100
x
X vs Y, facet by Z

WHY VISUALIZATIONS ?

“Visualizations can surprise you”

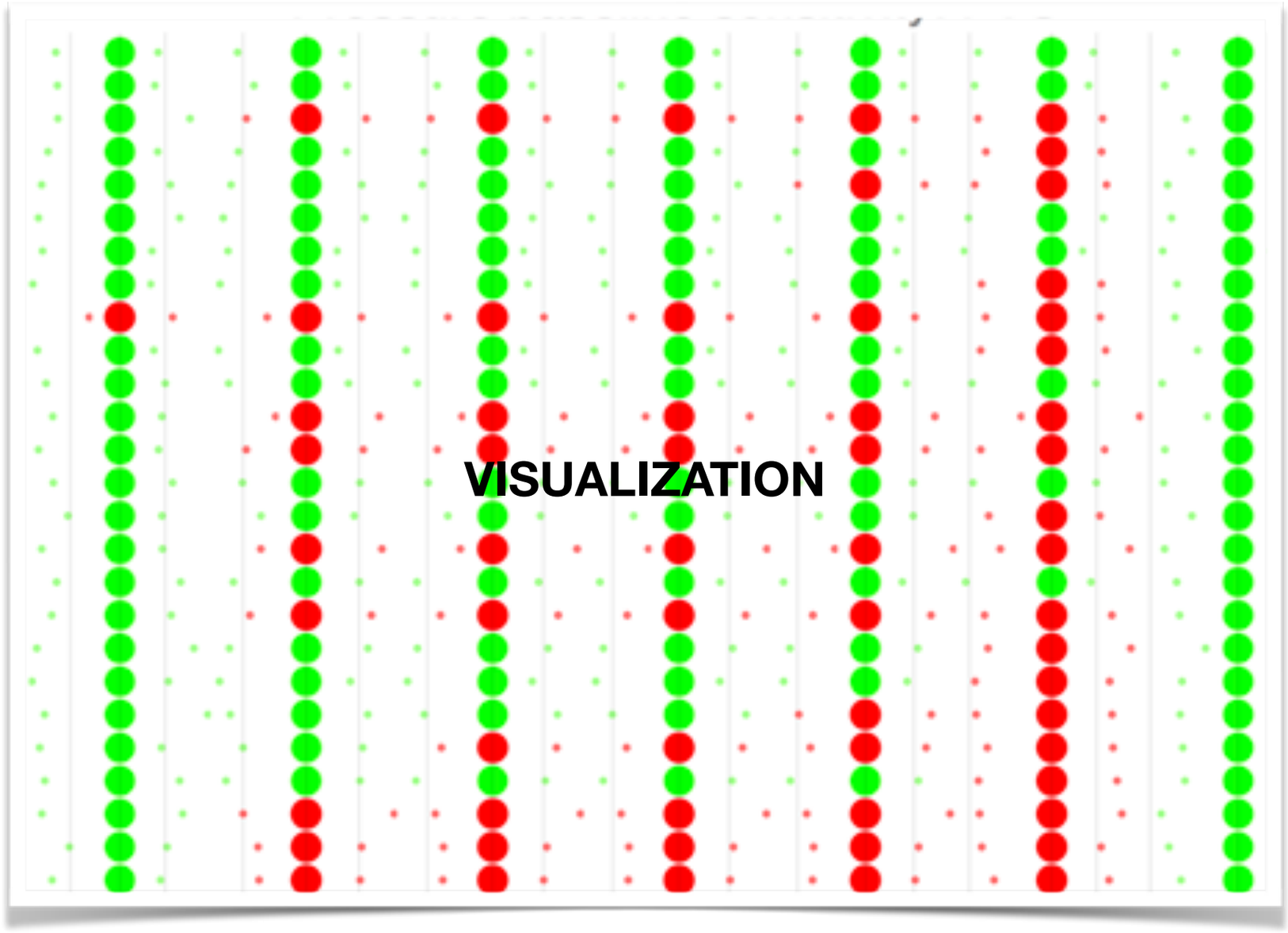
- Dr. Hadley Wickham

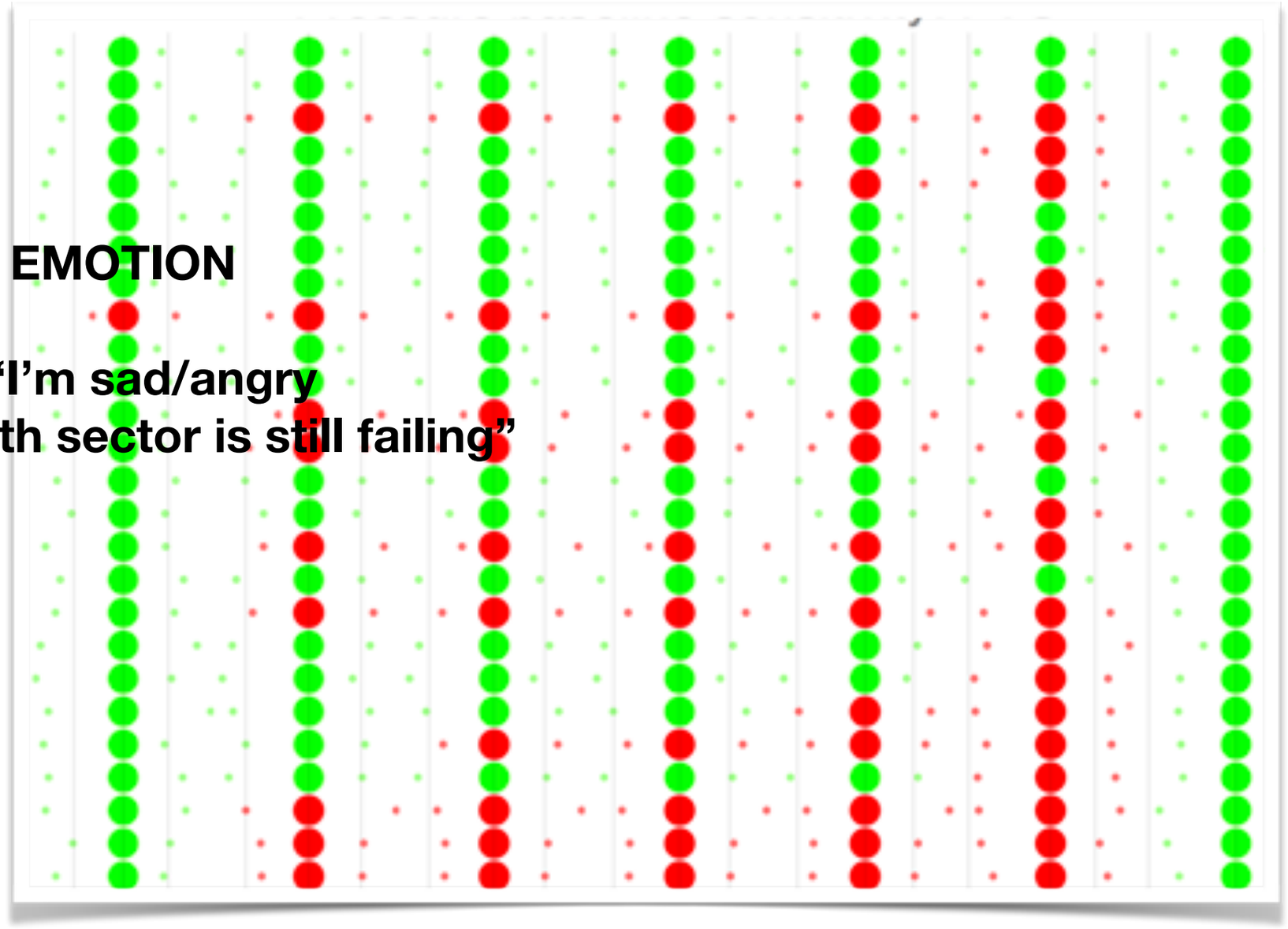
Data -> Visualizations -> Emotions -> Decisions -> Actions

- Dr. Anand Lakshmanan

f01	4.1749191781406	0.15455894200499998	3.702082253813509	3.280027722979433	0.06741096828976234
f02	3.8395846340118	0.16476975431800023	4.2913432056800405	3.4786189677219115	0.11918208935102026
f03	1.6383606029726001	0.12677683505100013	7.738030005175871	1.3802801553772217	0.06128564571183692
f04	1.7909154464608332	0.07022682691699988	3.9212809882108366	1.5467582715516506	0.05859430013147149
f05	2.8946250018836	0.14400622755699954	4.974952799180942	2.202101132380544	0.04476651230889628
f06	1.8232893974498	0.10271569613099985	5.633537729921883	1.3005688459565456	0.06072515412097301
f07	1.8269532432624	0.04861481164499981	2.6609773306616256	1.2880870250891847	0.01802114871234939
f08	1.7846759221886	0.07439246663699972	4.168401988959994	1.2017291189882398	0.0431324494202856
f09	2.7434286624924	0.14897372429900013	5.430202225985996	2.282360210752601	0.07459418956779329
f10	2.3481929137021997	0.06024658855000009	2.565657540249294	2.1022338325252523	0.07620932107805611
f11	2.763768398908	0.2200342533949997	7.961385385328887	2.100227881564459	0.07059657795876229
f12	5.1264973542400005	0.16337622368199956	3.1868976494618706	4.243360128845416	0.10574556713767347
f13	3.0175604929284	0.10261981958400046	3.4007543452563156	2.3629857759170148	0.07298021584006875
f14	3.0461130895601665	0.14378295911300043	4.[illegible]605543	2.370270702073574	0.040570802030806874
f15	3.5474003959963336	0.21505123510700042	6.062220530552782	3.0519798239632143	0.08498396306072609
f16	2.9060571990048	0.07706928671399949	2.6520223600689077	2.306750350700985	0.04256490357557219
f17	2.9414247275436	0.04083499071499963	1.3882725038862767	2.754675361006785	0.028036122177582
f18	4.5720515996628	0.16308494399600004	3.5669970130701922	3.99762811177299	0.14382704039224903
f19	3.6043237033798	0.08456529803999979	2.346218181255538	2.9669194236092267	0.018778781117449128
f20	5.331176432757	0.06786739876700043	1.273028563639246	4.3972488918881165	0.03339472081987793
f21	1.4286951577528	0.027021461506000044	1.8913384957853572	1.2265395508091264	0.04688161211614039
f22	2.0609493811754	0.09454002821899987	4.587207676351655	1.6971471060275023	0.06034120474063309
f23	4.272702461729001	0.09757922061500057	2.283782254650935	3.348270191763526	0.04886480399686155
f24	2.8776615701505	0.10902804545699984	3.7887723347293316	1.9573265997582214	0.08051001243593348
f25	5.833884322186	0.1368772414889996	2.34624538180268	5.393706337002649	0.09804550343722696
f26	4.8300593608878	0.09359876356499974	1.9378387835754387	4.048524432599245	0.06223400797501277

DATA


VISUALIZATION



EMOTION

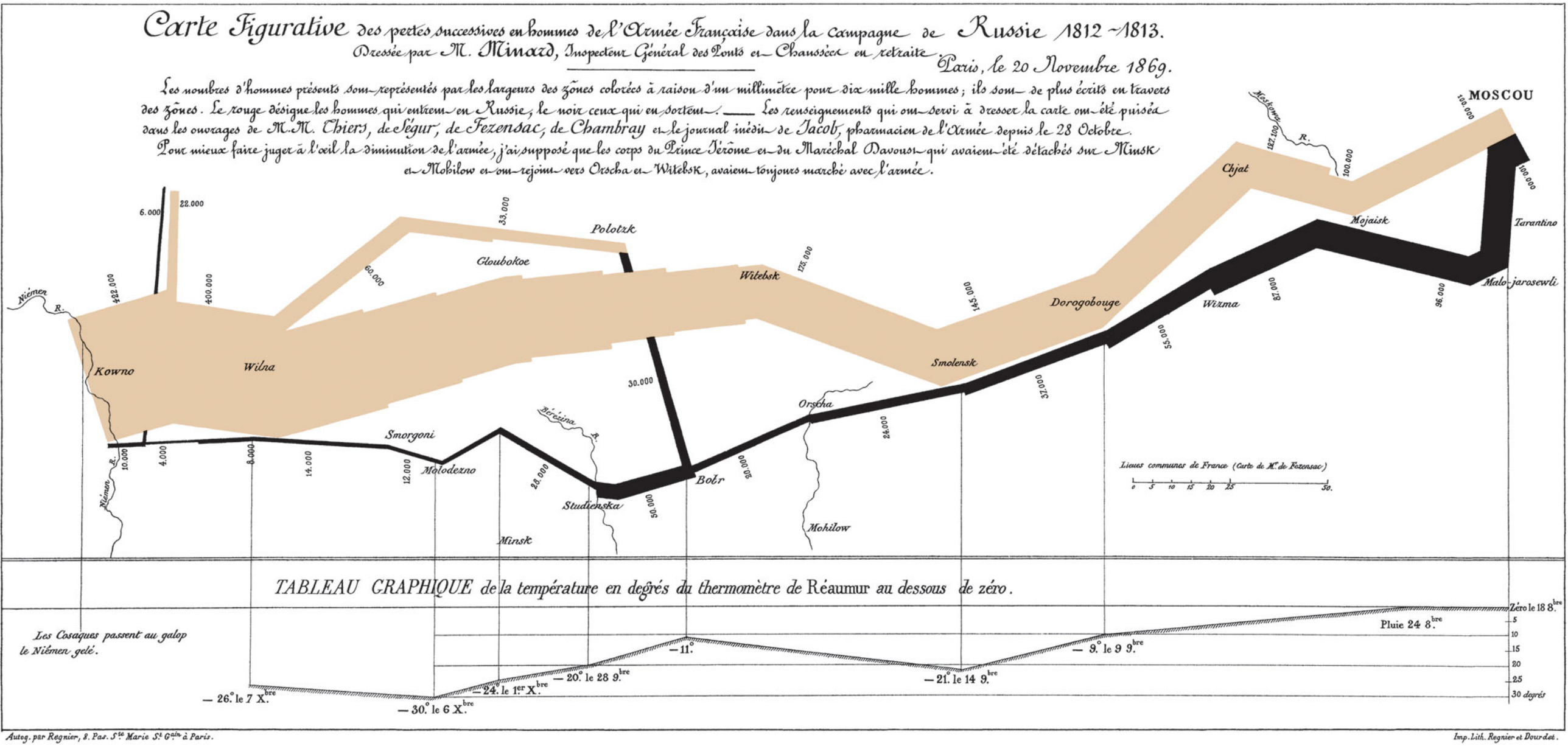
“I’m sad/angry
that the 9th sector is still failing”

DECISION & ACTION

“Allocate 10% more budget than last year”

“Change leadership”

Minard's Map



"Charles Minard's map of Napoleon's disastrous Russian campaign of 1812. The graphic is notable for its representation in two dimensions of six types of data: the number of Napoleon's troops; distance; temperature; the latitude and longitude; direction of travel; and location relative to specific dates"

Source: Wikimedia

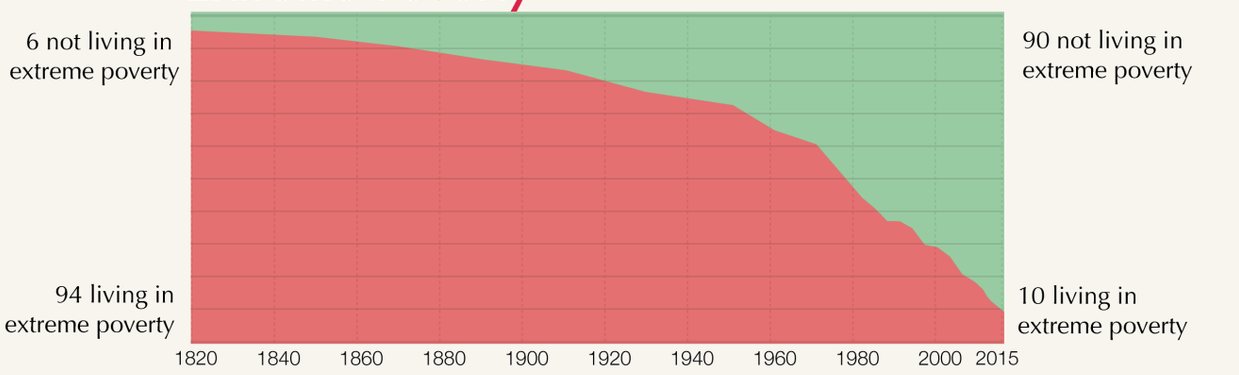
SIRPI
Visualize and Decide

INSOFE
Inspire...Educate...Transform.

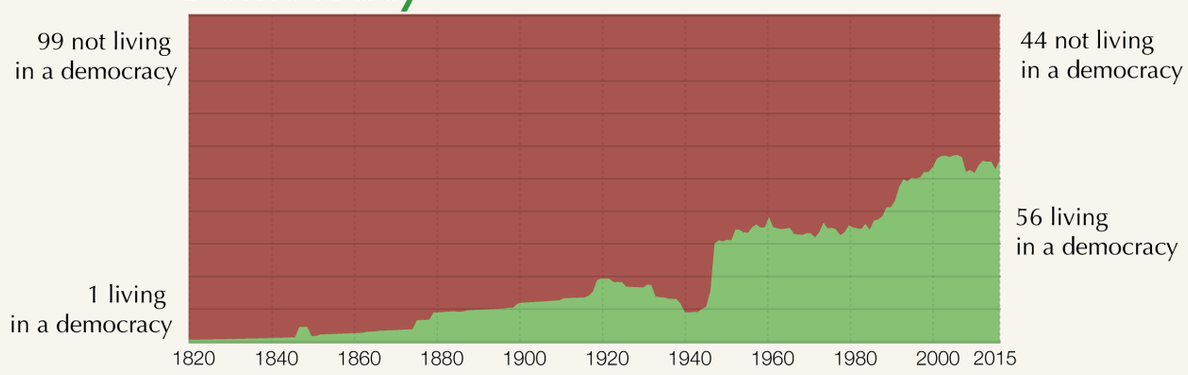
Our World in Data

The World as 100 People
over the last two centuries

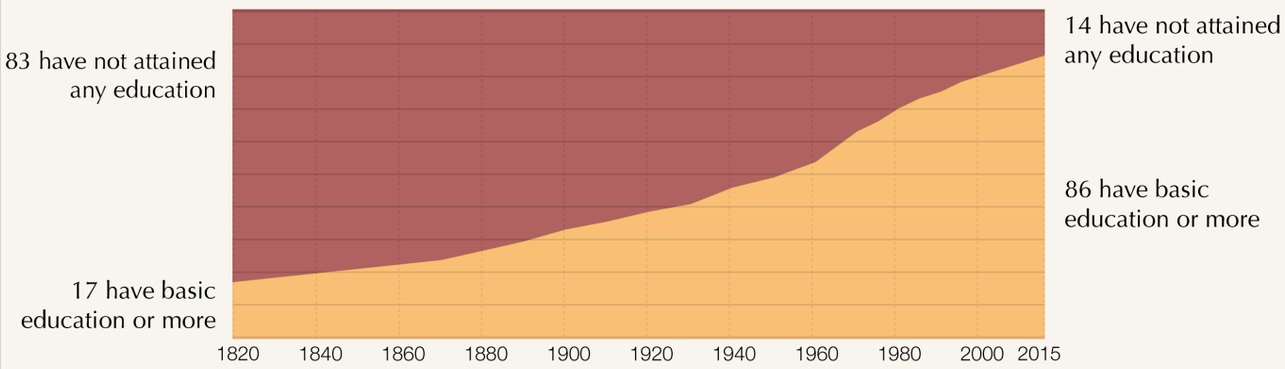
Extreme Poverty



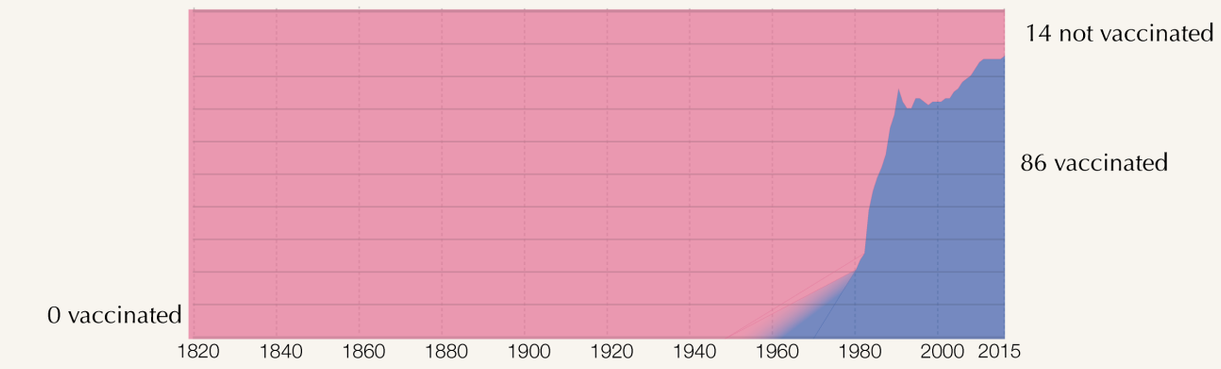
Democracy



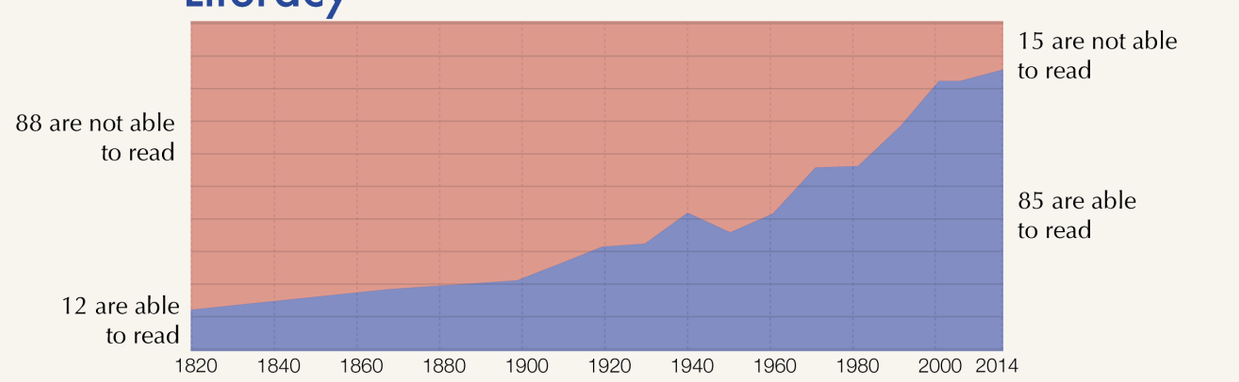
Basic Education



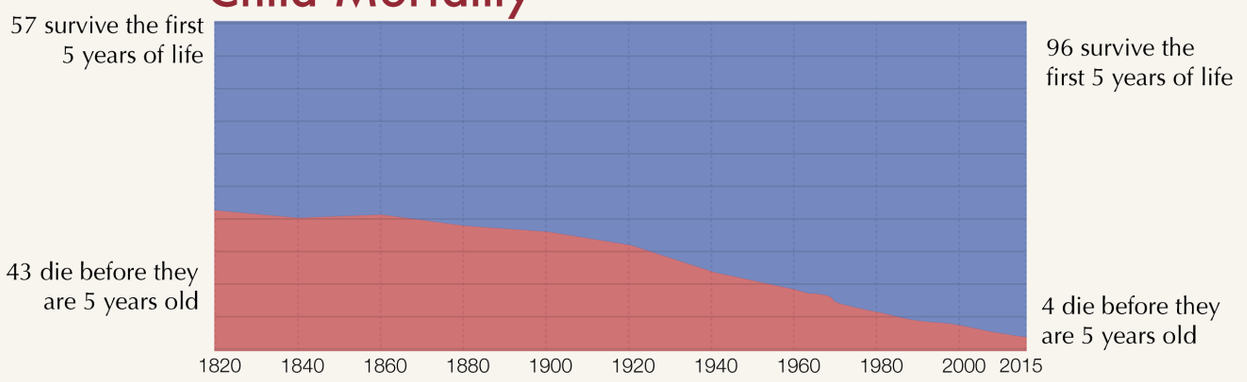
Vaccination against diphtheria, pertussis (whooping cough), and tetanus



Literacy

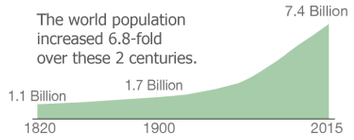


Child Mortality



Data sources:

Extreme Poverty: Bourguignon & Morrison (2002) up to 1970 – World Bank 1981 and later (2015 is a projection).
Vaccination: WHO (Global data are available for 1980 to 2015 – the DPT3 vaccination was licenced in 1949)
Education: OECD for the period 1820 to 1960. IIASA for the time thereafter.
Literacy: OECD for the period 1820 to 1990. UNESCO for 2004 and later.
Democracy: Politiy IV index (own calcluation of global population share)
Colonialism: Wimmer and Min (own calcluation of global population share)
Continent: HYDE database
Child mortality: up to 1960 own caluclations based on Gapminder; World Bank thereafter



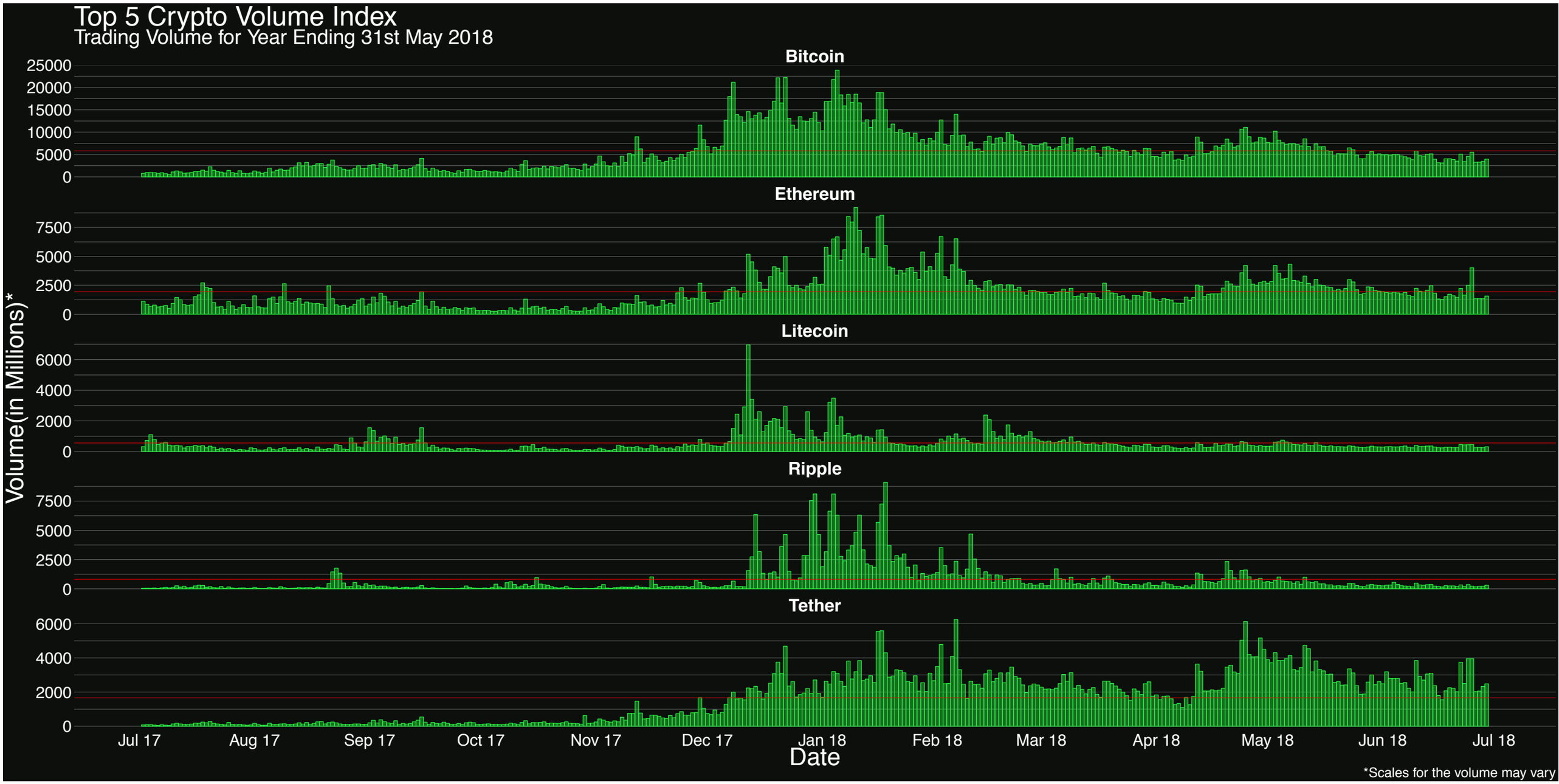
All these visualizations are from OurWorldInData.org an online publication that presents the empirical evidence on how the world is changing.

Source: Wikimedia

SIRPI
Visualize and Decide

INSOFE
Inspire...Educate...Transform.

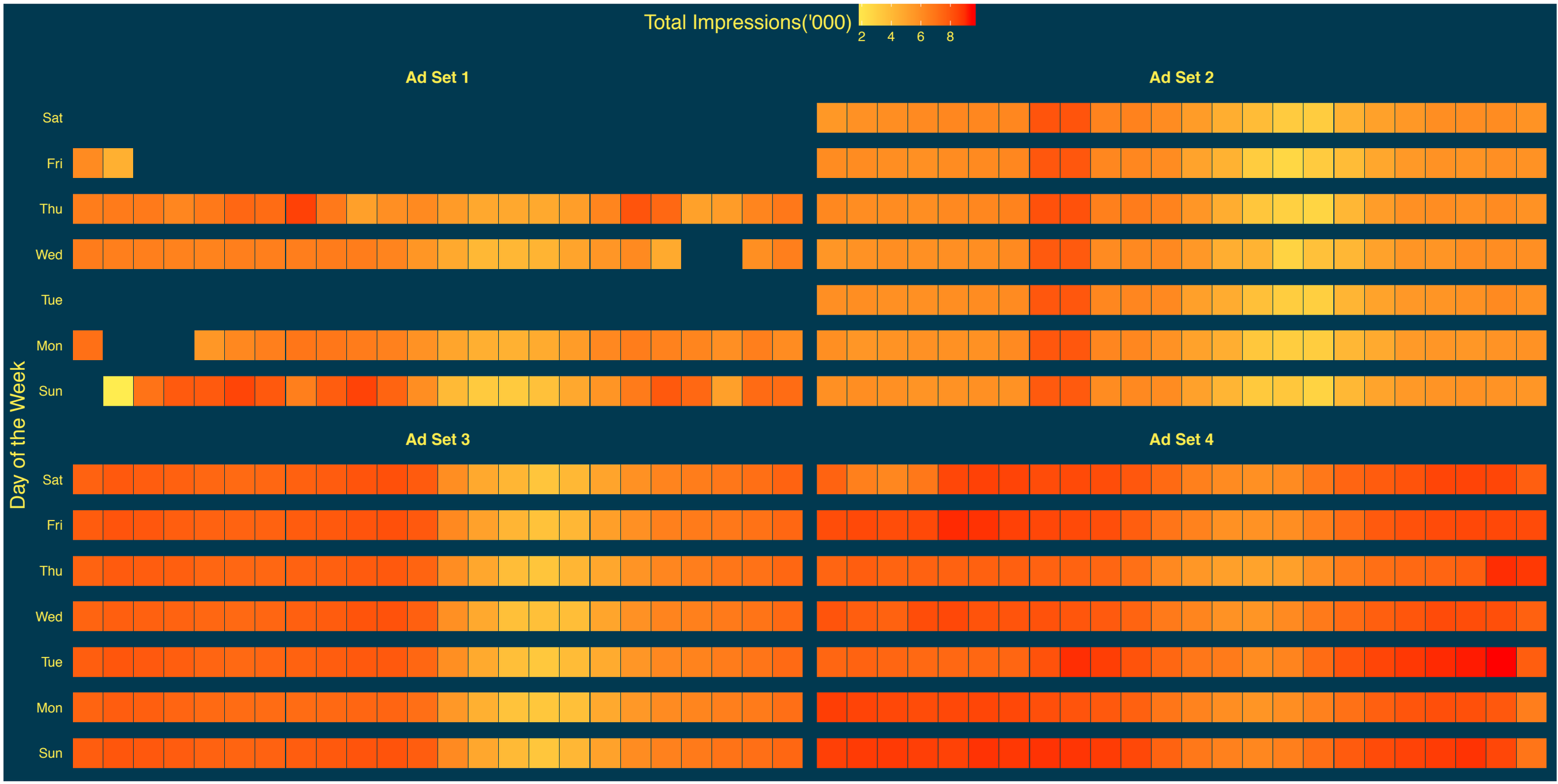
Visualizing Cryptocurrencies



SIRPI
Visualize and Decide

INSOFE
Inspire...Educate...Transform.

AdROI-T

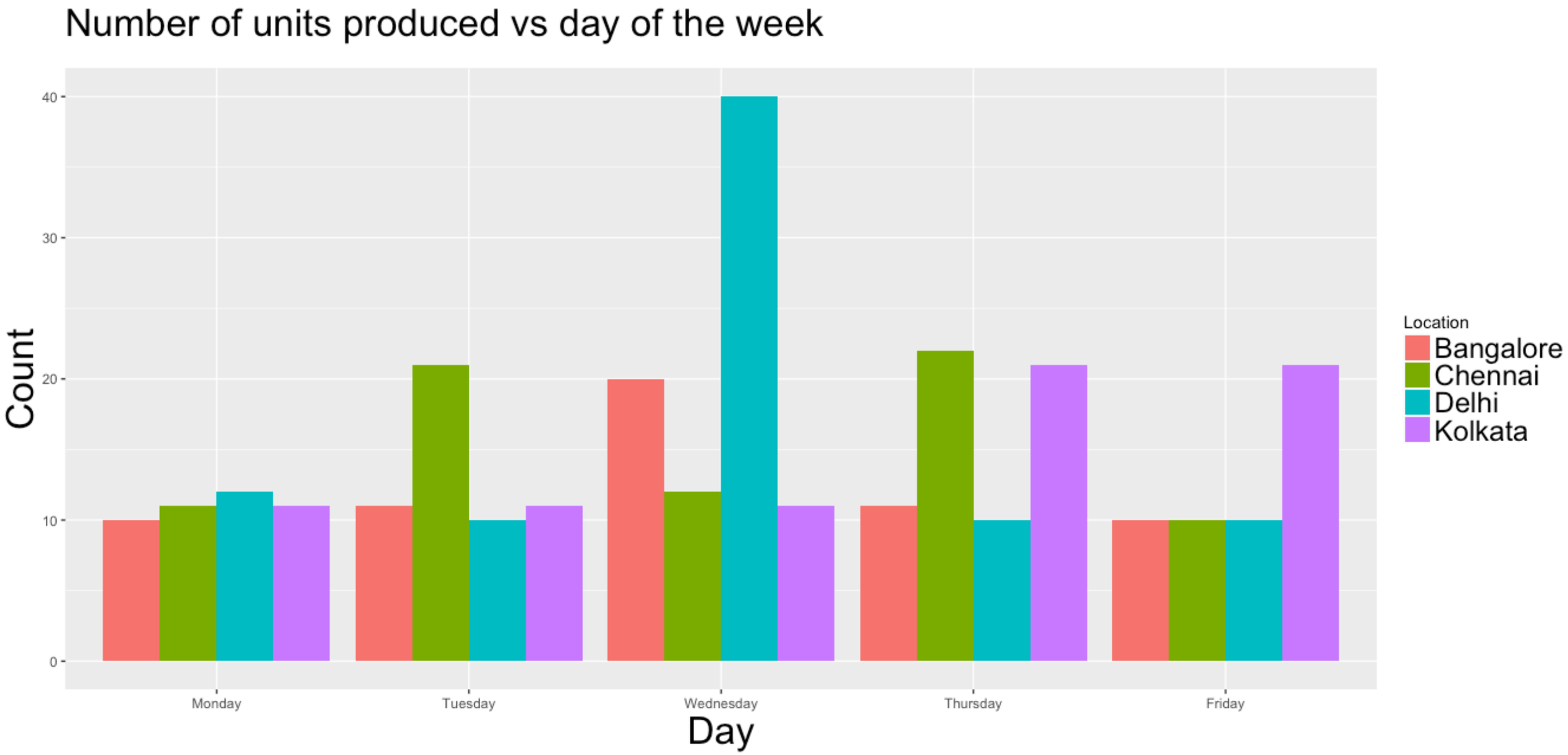


Attention

Use clear + bold + short visual to draw attention
Example : Traffic speed limit

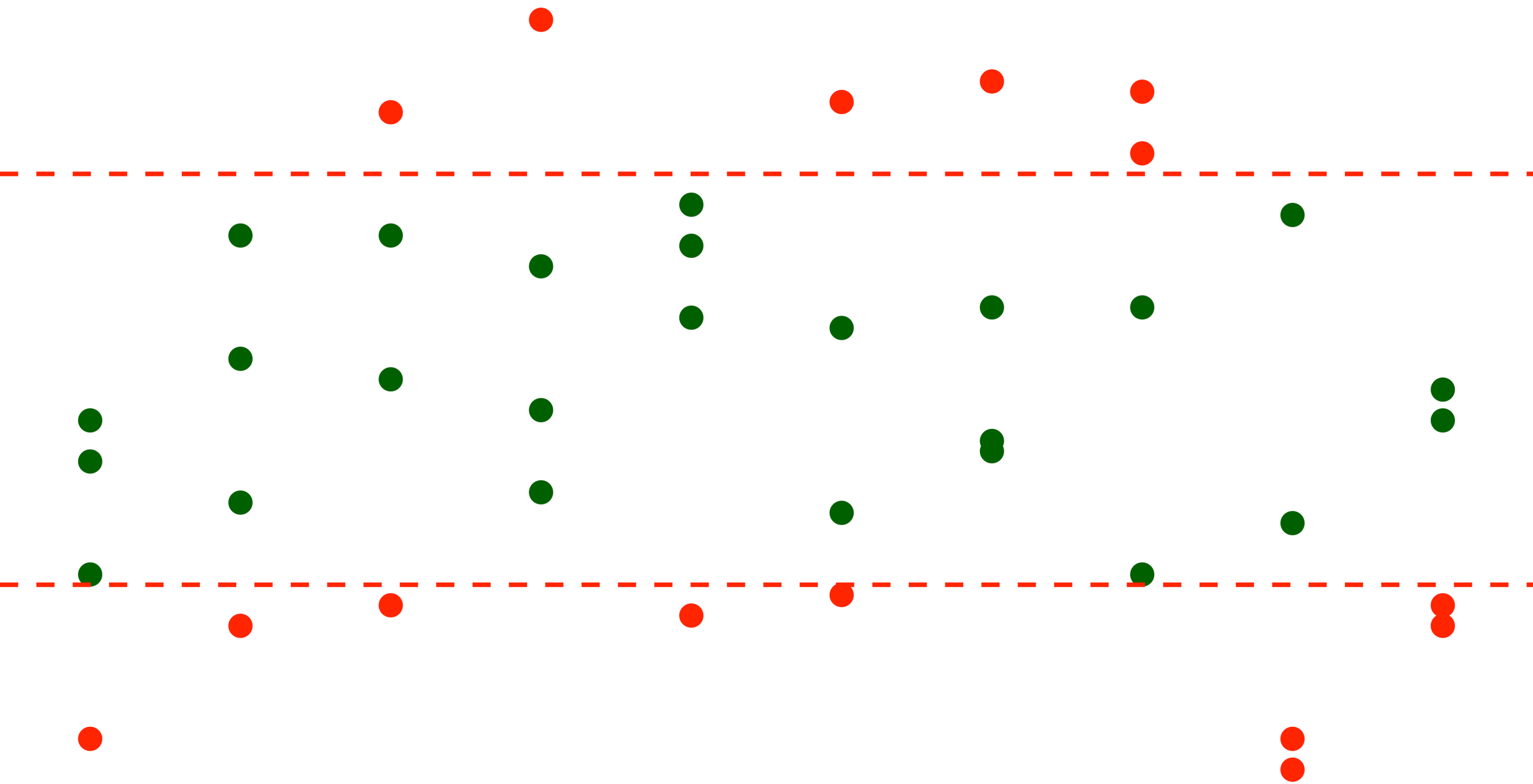
Source: Wikipedia

Comparison

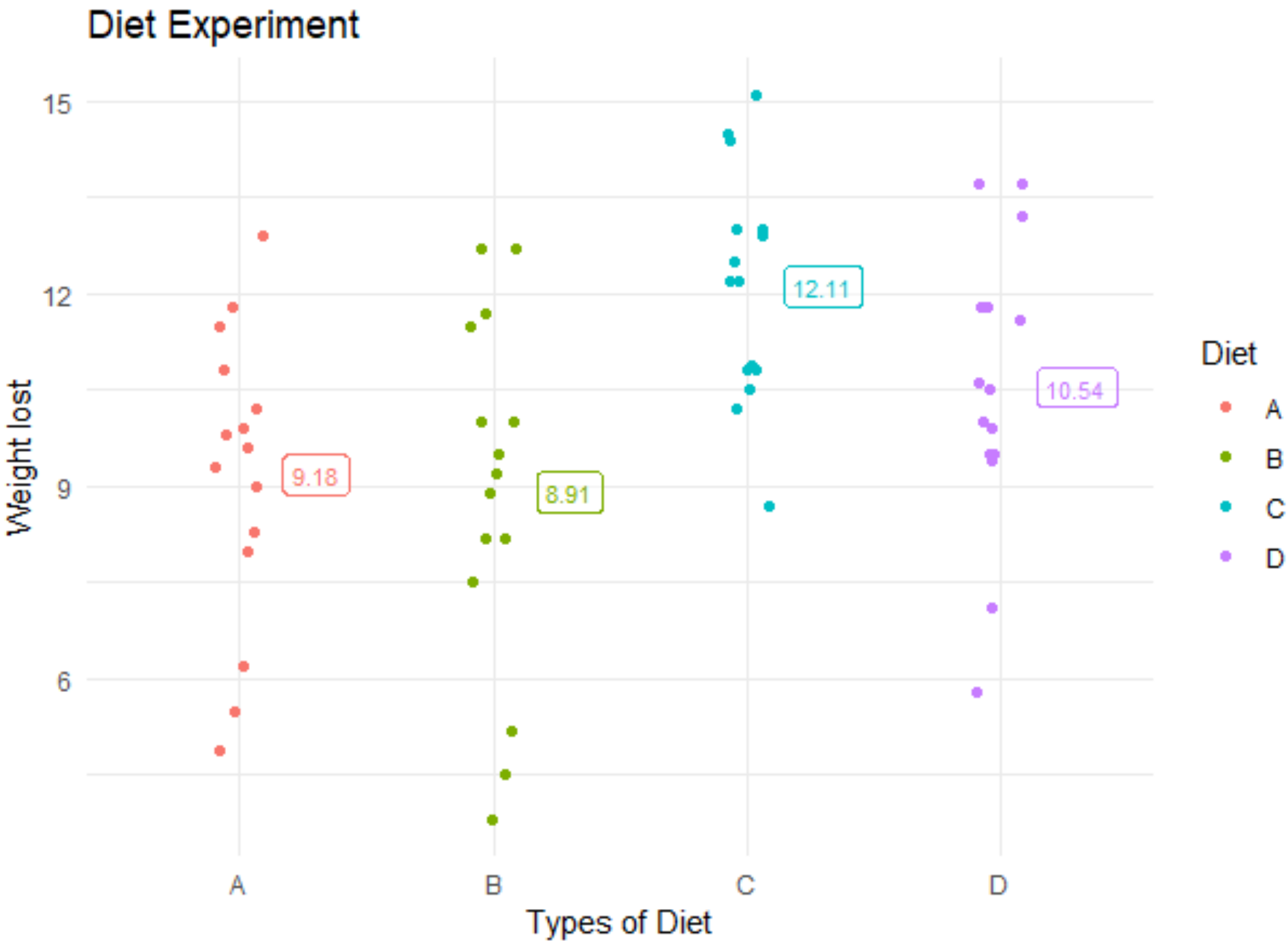


Use same axes limits to compare within and across parameters

Quality Control Pass / Fail



Group Analysis



Reference

1 H Hydrogen 1.008																	2 He Helium 4.003
3 Li Lithium 6.94	4 Be Beryllium 9.012											5 B Boron 10.81	6 C Carbon 12.011	7 N Nitrogen 14.007	8 O Oxygen 15.999	9 F Fluorine 18.998	10 Ne Neon 20.180
11 Na Sodium 22.990	12 Mg Magnesium 24.305											13 Al Aluminum 26.982	14 Si Silicon 28.085	15 P Phosphorus 30.974	16 S Sulfur 32.06	17 Cl Chlorine 35.45	18 Ar Argon 39.948
19 K Potassium 39.098	20 Ca Calcium 40.078	21 Sc Scandium 44.956	22 Ti Titanium 47.867	23 V Vanadium 50.942	24 Cr Chromium 51.996	25 Mn Manganese 54.938	26 Fe Iron 55.845	27 Co Cobalt 58.933	28 Ni Nickel 58.693	29 Cu Copper 63.546	30 Zn Zinc 65.38	31 Ga Gallium 69.723	32 Ge Germanium 72.630	33 As Arsenic 74.922	34 Se Selenium 78.97	35 Br Bromine 79.904	36 Kr Krypton 83.798
37 Rb Rubidium 85.468	38 Sr Strontium 87.62	39 Y Yttrium 88.906	40 Zr Zirconium 91.224	41 Nb Niobium 92.906	42 Mo Molybdenum 95.95	43 Tc Technetium [97]	44 Ru Ruthenium 101.07	45 Rh Rhodium 102.906	46 Pd Palladium 106.42	47 Ag Silver 107.868	48 Cd Cadmium 112.414	49 In Indium 114.818	50 Sn Tin 118.710	51 Sb Antimony 121.760	53 Te Tellurium 127.60	53 I Iodine 126.904	54 Xe Xenon 131.293
55 Cs Cesium 132.905	56 Ba Barium 137.327	* 57 - 70	71 Lu Lutetium 174.967	72 Hf Hafnium 178.49	73 Ta Tantalum 180.948	74 W Tungsten 183.84	75 Re Rhenium 186.207	76 Os Osmium 190.23	78 Ir Iridium 192.217	79 Pt Platinum 195.084	80 Au Gold 196.997	81 Hg Mercury 200.592	81 Tl Thallium 204.38	82 Pb Lead 207.2	83 Bi Bismuth 208.980	84 Po Polonium [209]	85 At Astatine [210]	86 Rn Radon [222]
87 Fr Francium [223]	88 Ra Radium [226]	** 89 - 102	103 Lr Lawrencium [262]	104 Rf Rutherfordium [267]	105 Db Dubnium [270]	106 Sg Seaborgium [269]	107 Bh Bohrium [270]	108 Hs Hassium [270]	109 Mt Meitnerium [278]	110 Ds Darmstadtium [281]	111 Rg Roentgenium [281]	112 Cn Copernicium [285]	113 Nh Nihonium [286]	114 Fl Flerovium [289]	115 Mc Moscovium [289]	116 Lv Livermorium [293]	117 Ts Tennessine [293]	118 Og Oganesson [294]

*Lanthanide series	57 La Lanthanum 138.905	58 Ce Cerium 140.116	59 Pr Praseodymium 140.908	60 Nd Neodymium 144.242	61 Pm Promethium [145]	62 Sm Samarium 150.36	63 Eu Europium 151.964	64 Gd Gadolinium 157.25	65 Tb Terbium 158.925	66 Dy Dysprosium 162.500	67 Ho Holmium 164.930	68 Er Erbium 167.259	69 Tm Thulium 168.934	70 Yb Ytterbium 173.045
**Actinide series	89 Ac Actinium [227]	90 Th Thorium 232.038	91 Pa Protactinium 231.036	92 U Uranium 238.029	93 Np Neptunium [237]	94 Pu Plutonium [244]	95 Am Americium [243]	96 Cm Curium [247]	97 Bk Berkelium [247]	98 Cf Californium [251]	99 Es Einsteinium [252]	100 Fm Fermium [257]	101 Md Mendelevium [258]	102 No Nobelium [259]

Periodic table is a great example of visualization for reference.

Source: Wikimedia

SIRPI
Visualize and Decide

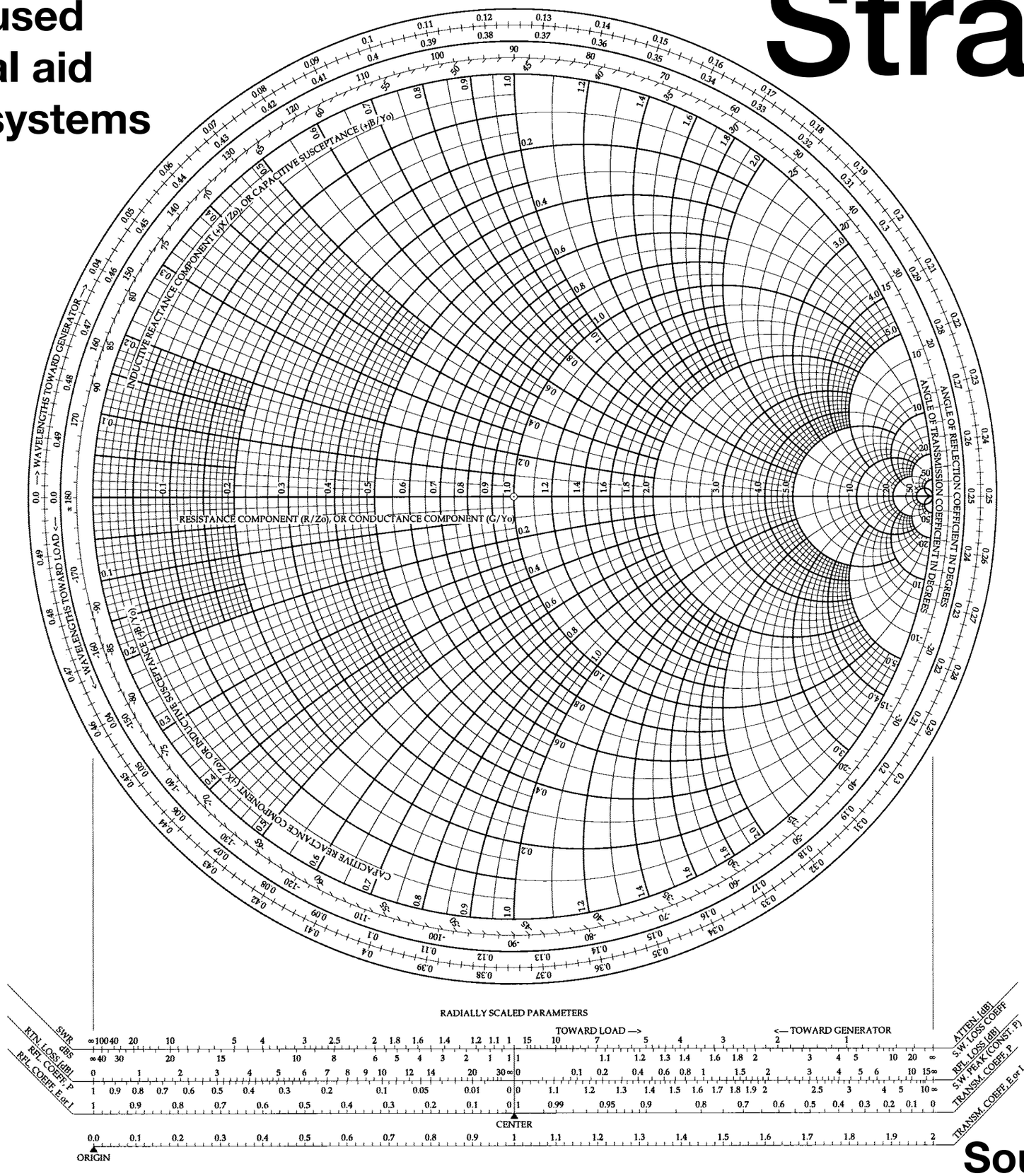
INSOFE
Inspire...Educate...Transform.

The Complete Smith Chart

Black Magic Design

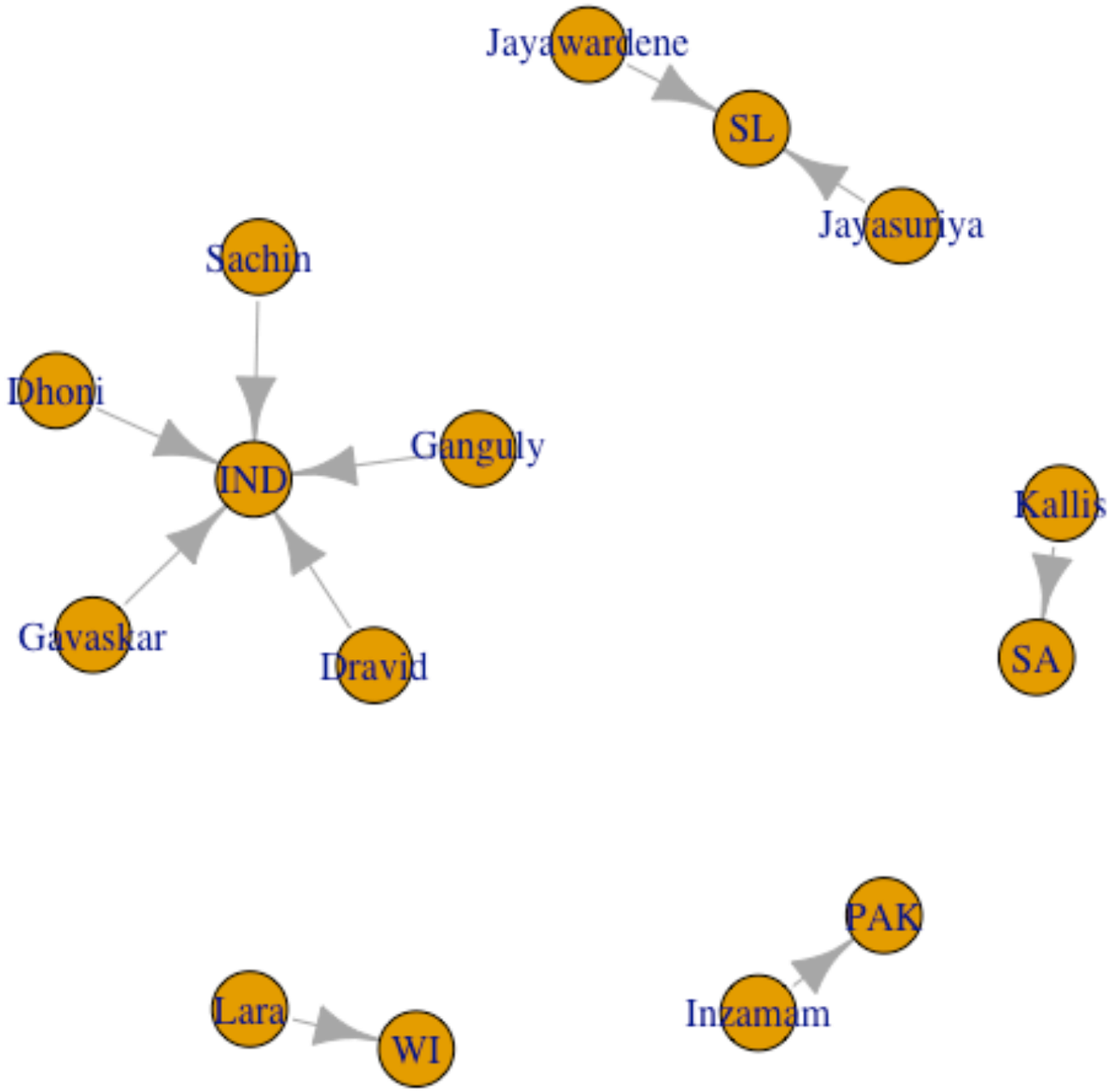
A smith chart is used as a mathematical aid to design wireless systems

Strategy



Source: Wikimedia

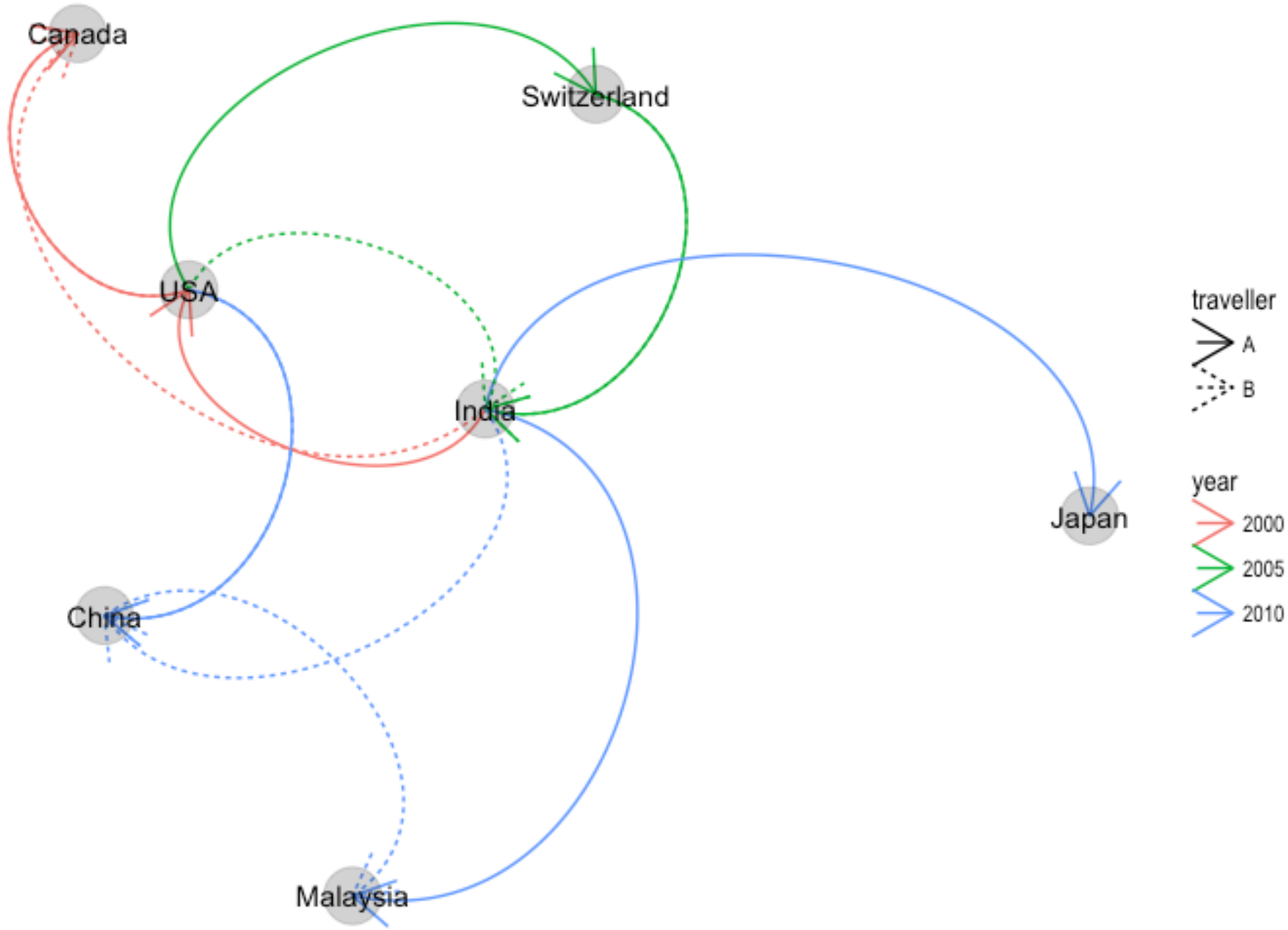
Network



Network Visualization Example

Source: Wikimedia

Network



Network Visualization Example

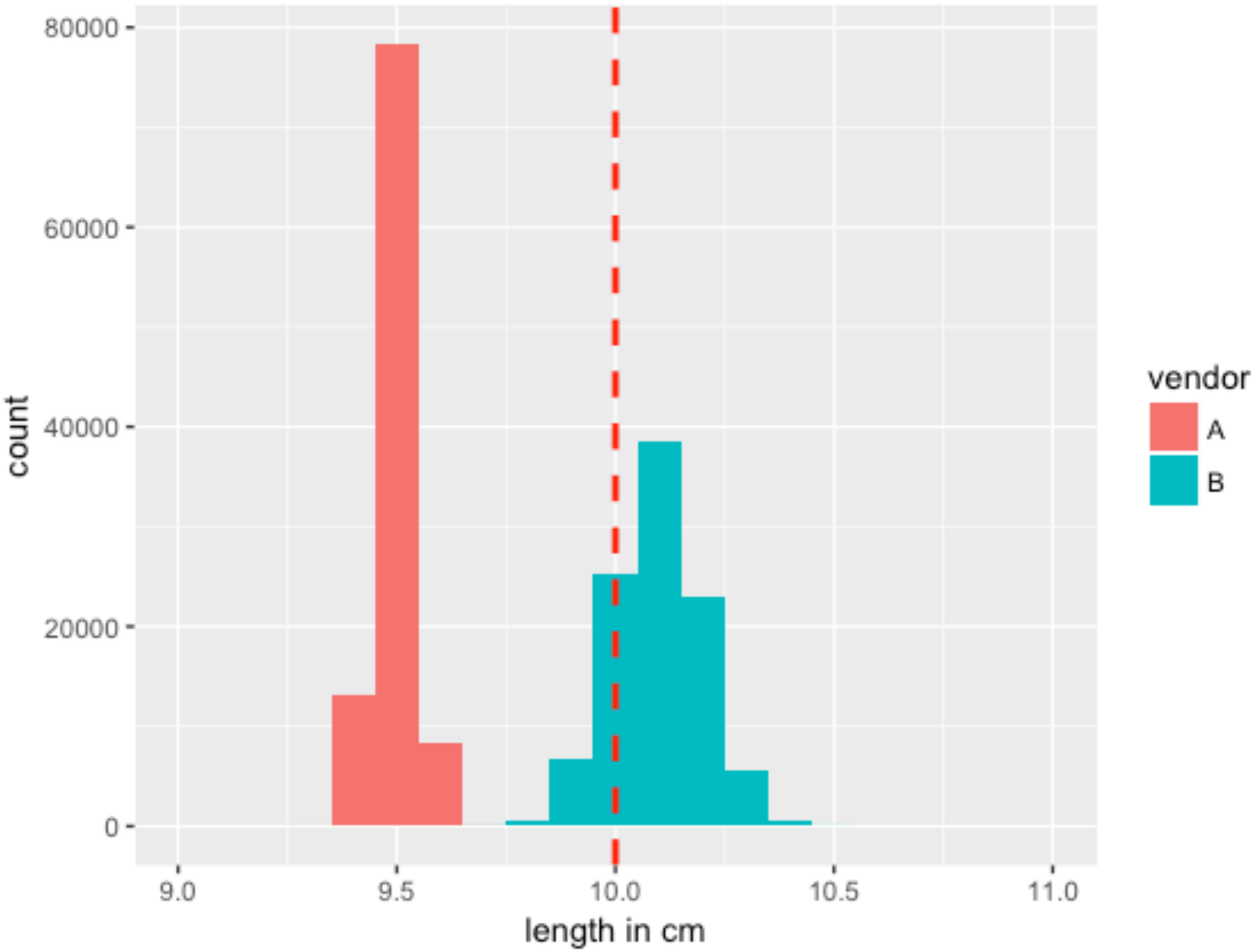
Source: Wikimedia

Case Study

Decision Challenge !

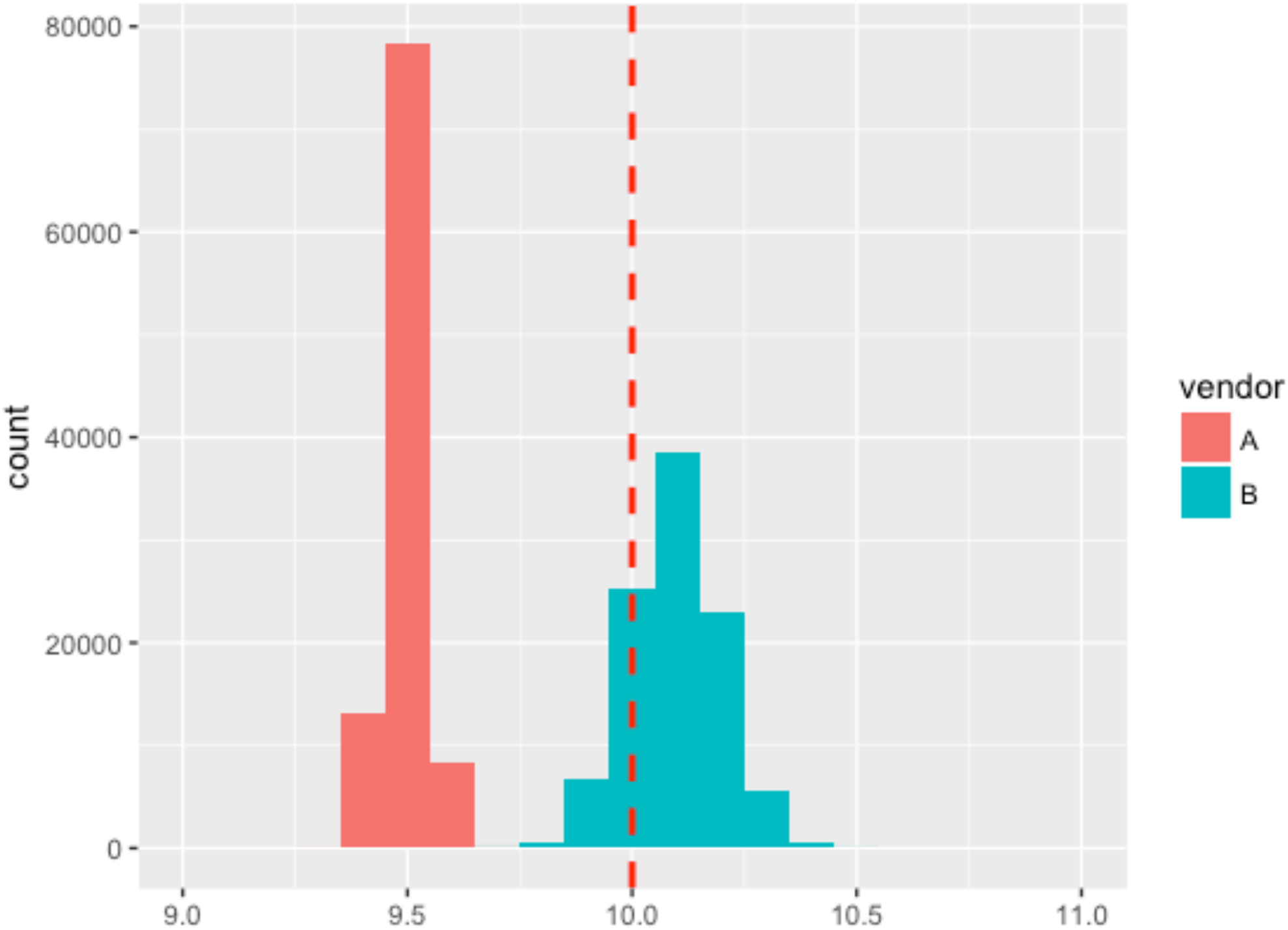
Vendor A and B each produced 1 lakh cables that are supposed to be 10 cm each.
Which vendor would you hire for manufacturing 50 lakh cables and why ?

Each vendor produces 1 lakh cables that are supposed to be 10 cm each. Which vendor would you hire for manufacturing 50 lakh cables and why ?



**Hire B for now,
choose bin around 10 cm,
reject rest of the samples !**

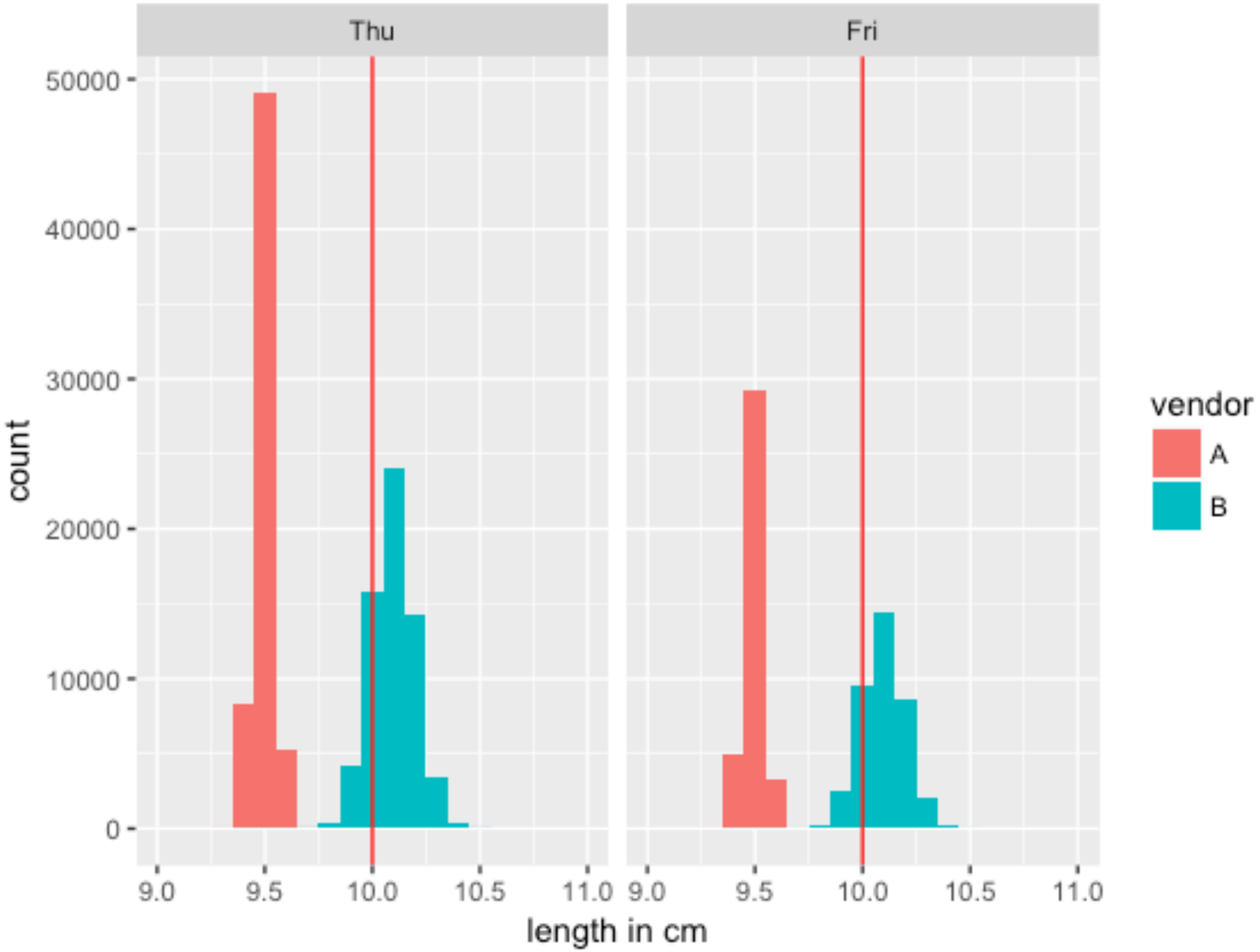
- Old paradigm : Use theoretical knowledge to solve problems
- New paradigm : Collect data, visualize data, propose solutions



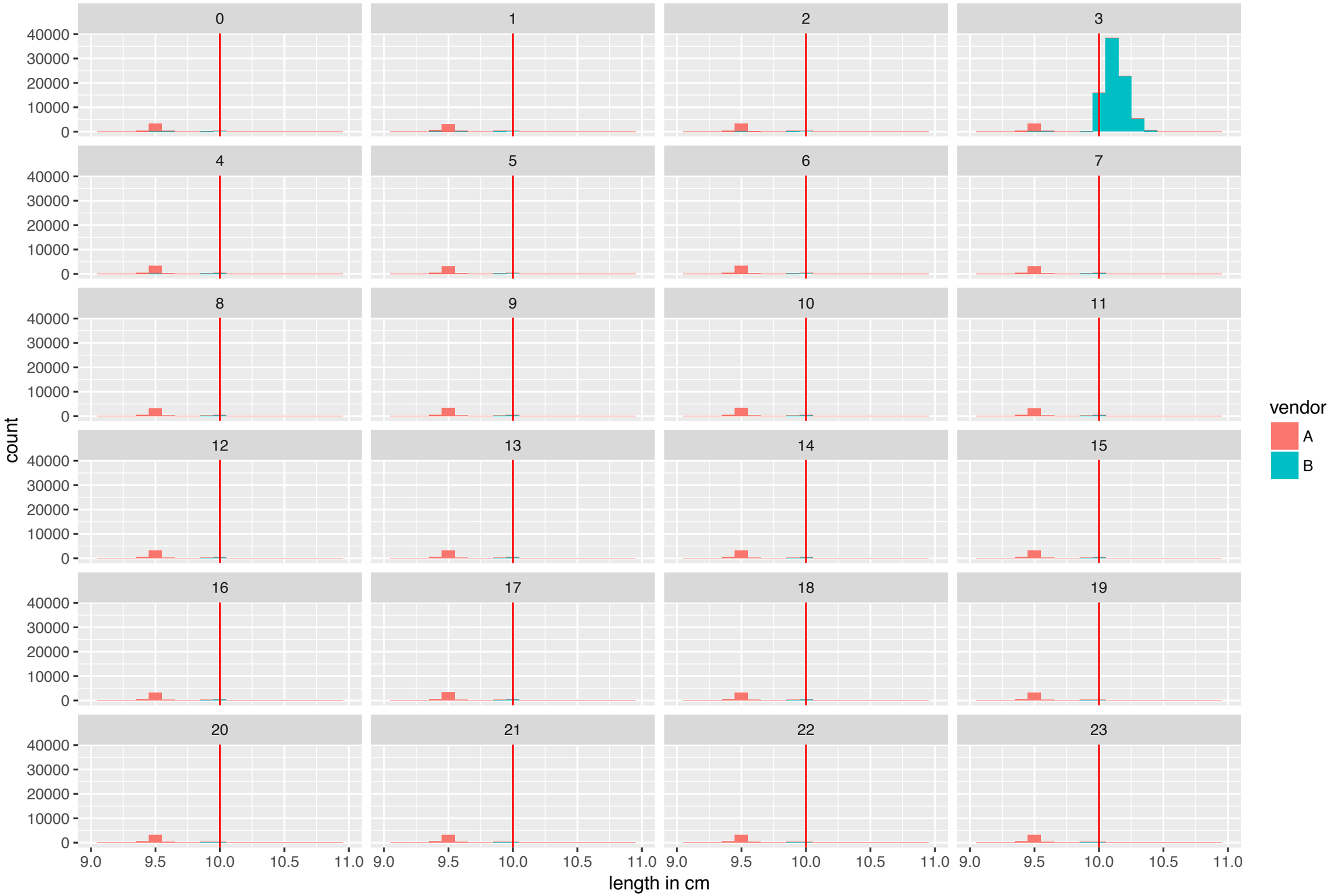
**A has long term potential:
low variance !**

Ask for timestamps

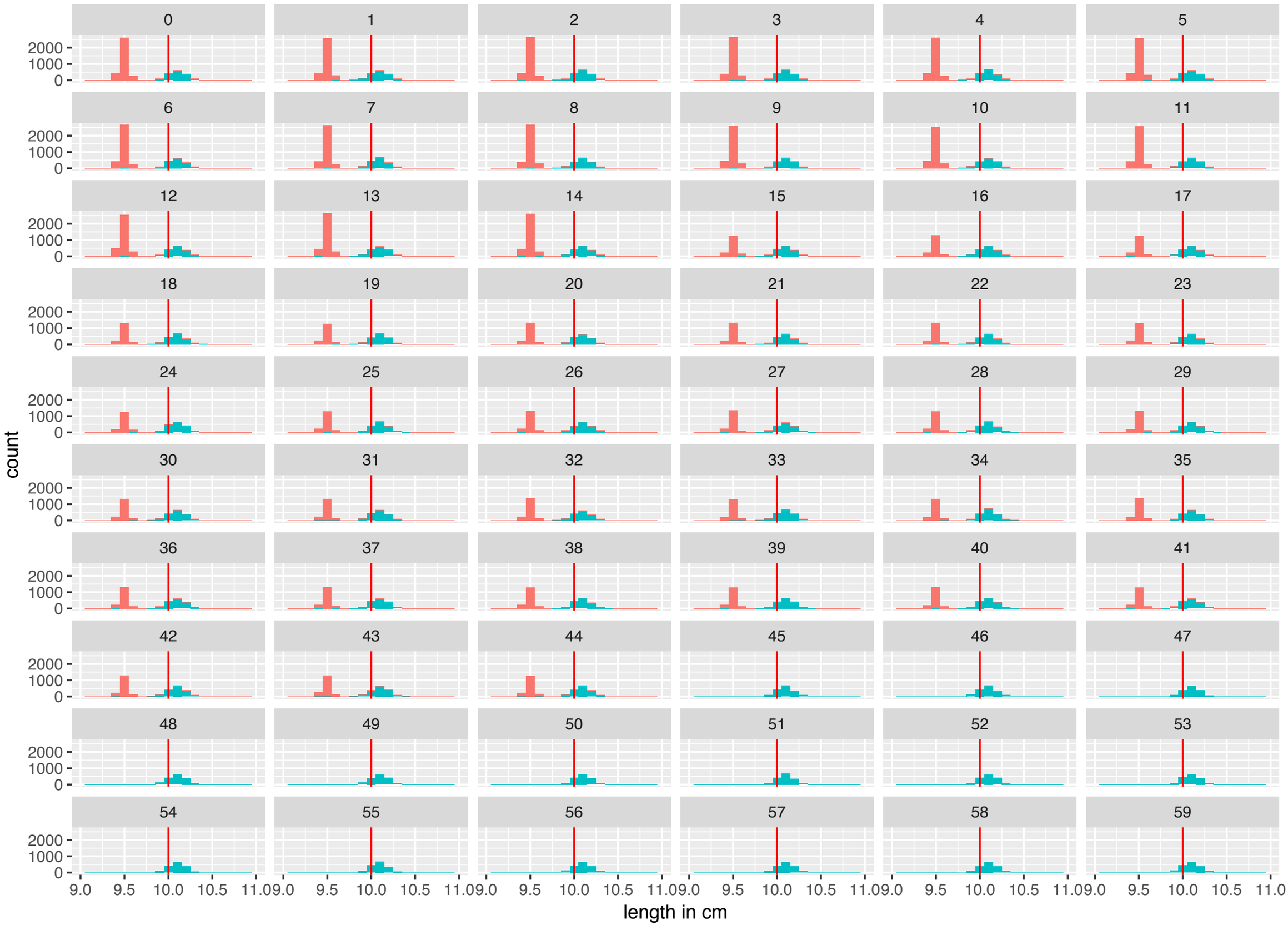
Vendor vs cable length



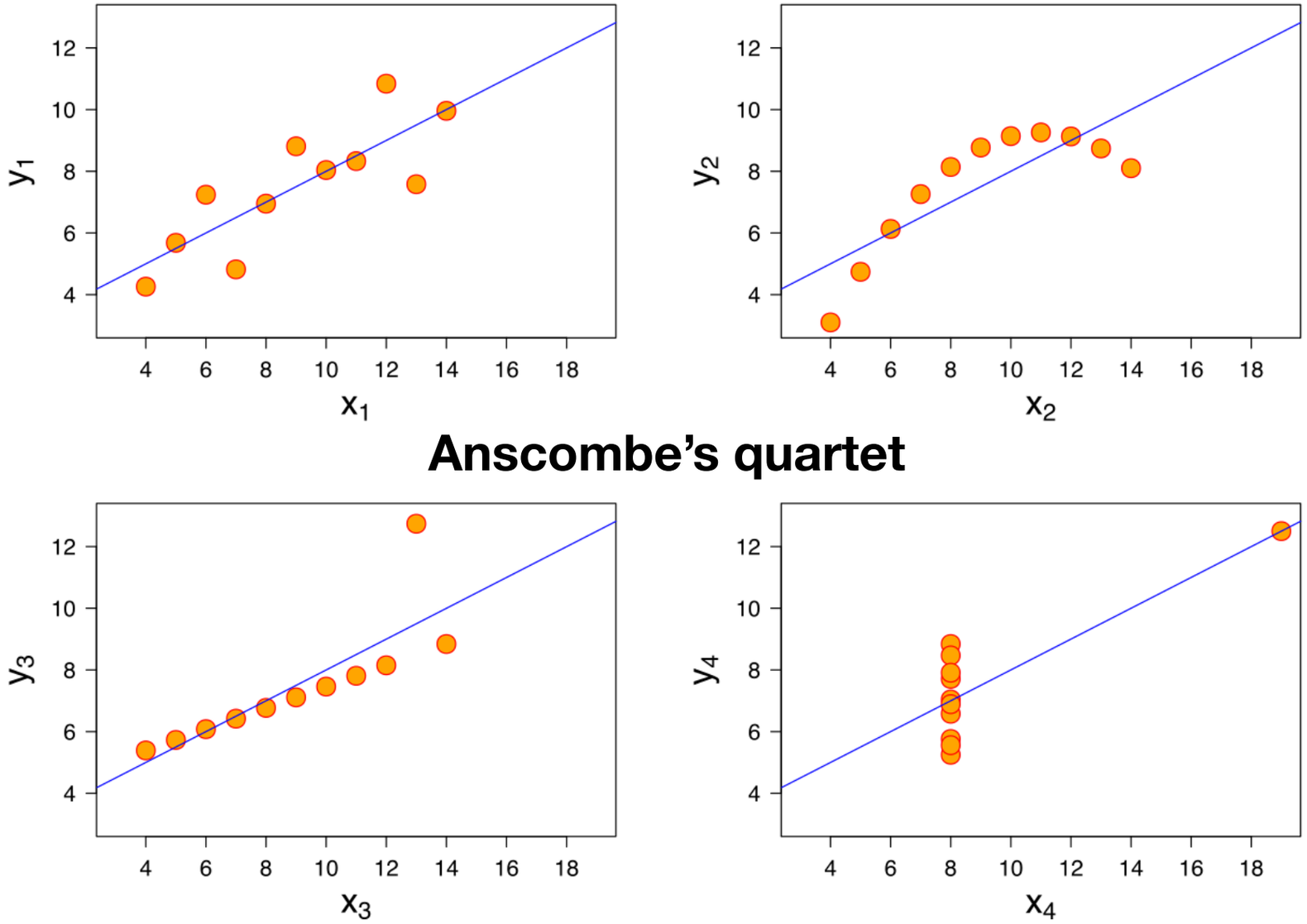
Vendor vs cable length (by hour of the day)



Vendor vs cable length (by minute of the hour)



Debunking Evidence

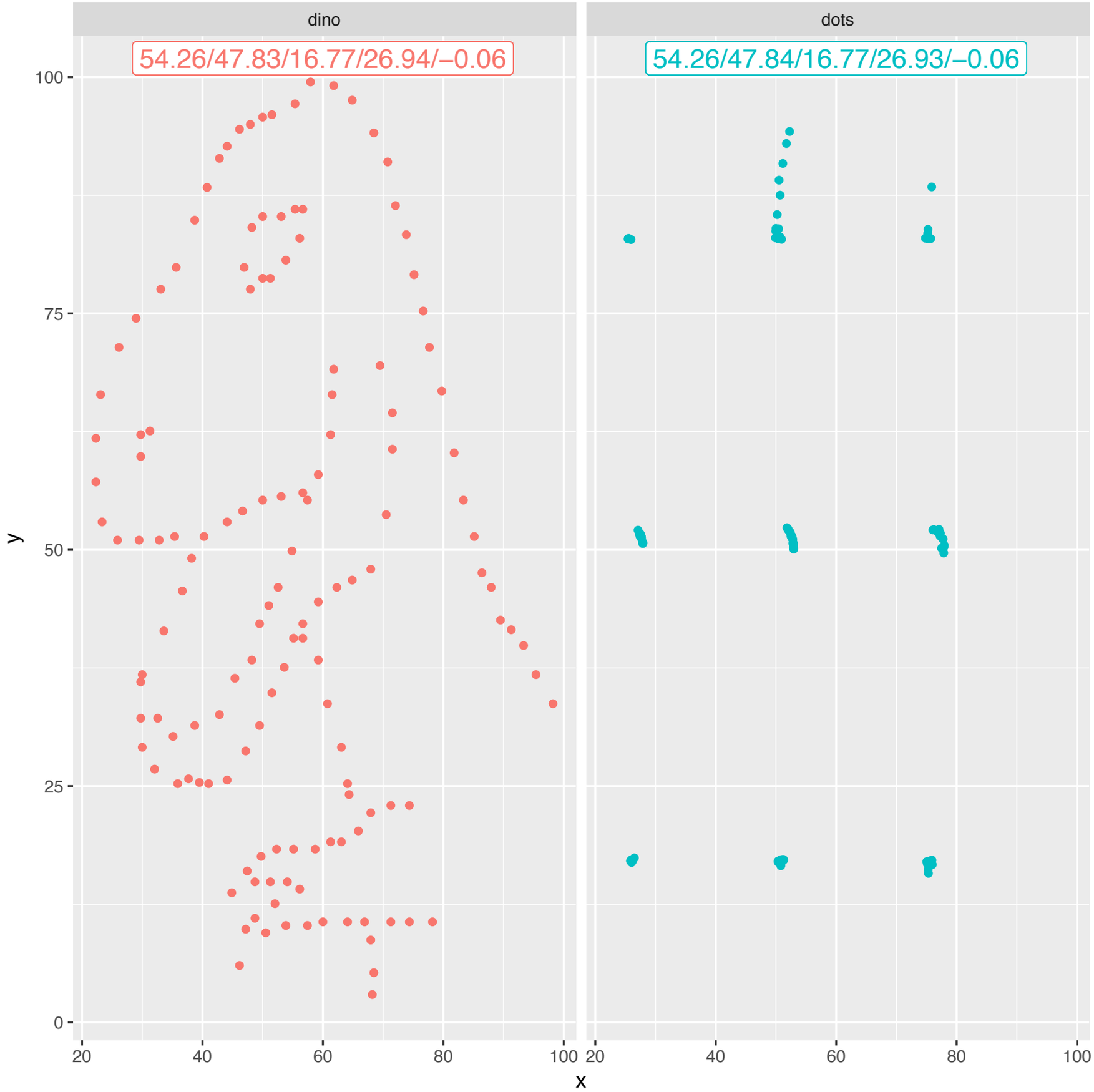


Anscombe's quartet

Ask for raw data in addition to stats and trend lines.

"**Anscombe's quartet** comprises four datasets that have nearly identical simple descriptive statistics, yet appear very different when graphed. Each dataset consists of eleven (*x*,*y*) points. They were constructed in 1973 by the statistician Francis Anscombe to demonstrate both the importance of graphing data before analyzing it and the effect of outliers on statistical properties."

Source: Wikipedia


dino
dots
54.26/47.83/16.77/26.94/–0.06
54.26/47.84/16.77/26.93/–0.06
100
75
50
25
0
y
20
40
60
80
100
x

SIRPI
Visualize and Decide

INSOFE
Inspire...Educate...Transform.

SCIENCE

Theory & Experiment

We have been conditioned to think theory drives experiment

Academia : “Do this experiment to verify that the results match theory”

Business : “This is how we’ve always done things in this company”

Scientific Method

- Hypothesis - Initial guess
- Experimentation
- Inference
- Formulate theory
- Validate theory

Scientific Method

- Hypothesis - Initial guess
- Experimentation
- Inference
- Formulate theory
- Validate theory

Experiment drives theory !

Mindset change is needed :

Experiment drives theory

Next generation academia :
My experimental data shows this is were theory fails, need modification to this theory.

Next generation business :
New data collected shows we need to focus *here* to improve sales

DATA

VR

API

ROBOTICS

WEB

AUDIO

VIDEO

TEXT

TIME SERIES

STATISTICS

DATA

DATA BASE

CSV

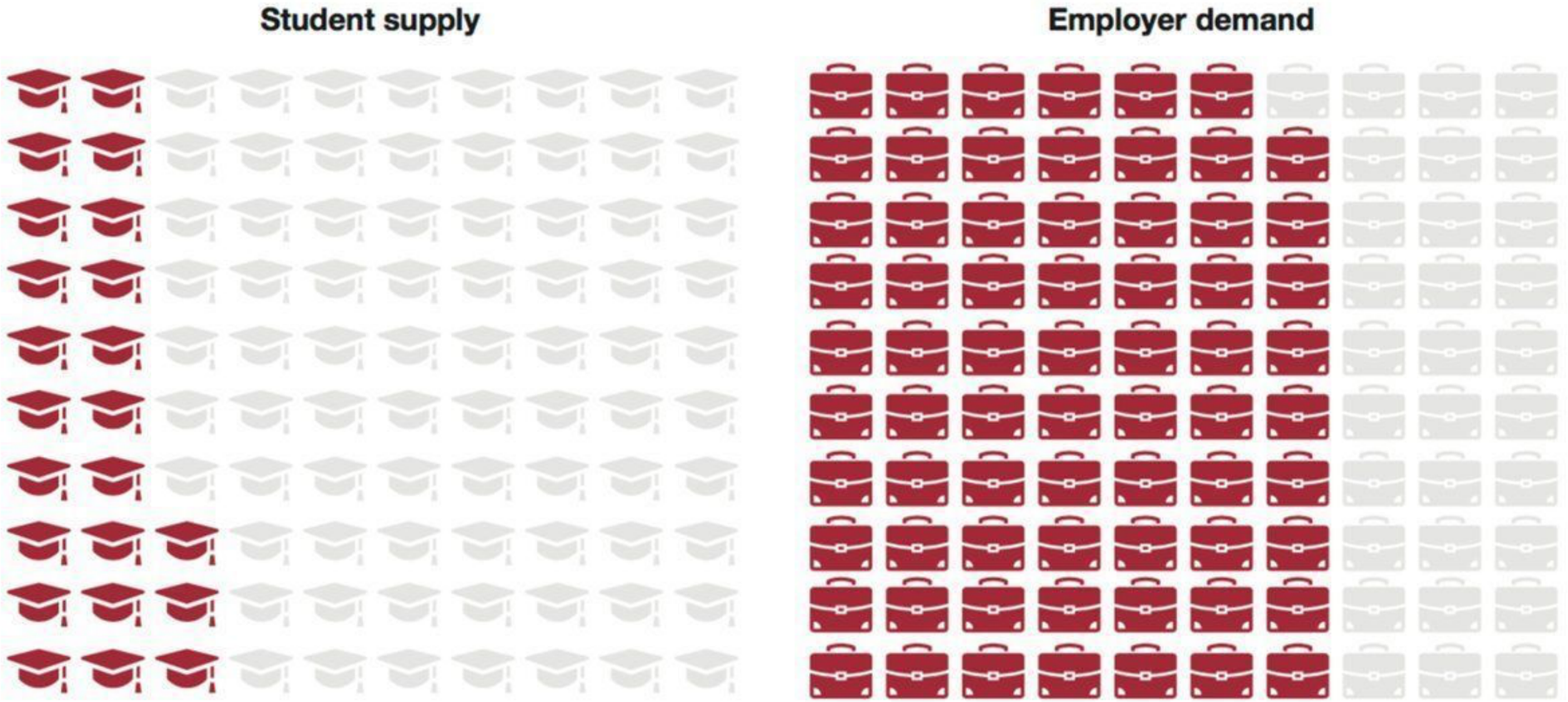
EXCEL

BLOCKCHAIN

AR

Figure 1: Data science and analytics skills, by 2021

The supply-demand challenge



23% of educators say all graduates will have data science and analytics skills

69% of employers say they will prefer job candidates with these skills over ones without

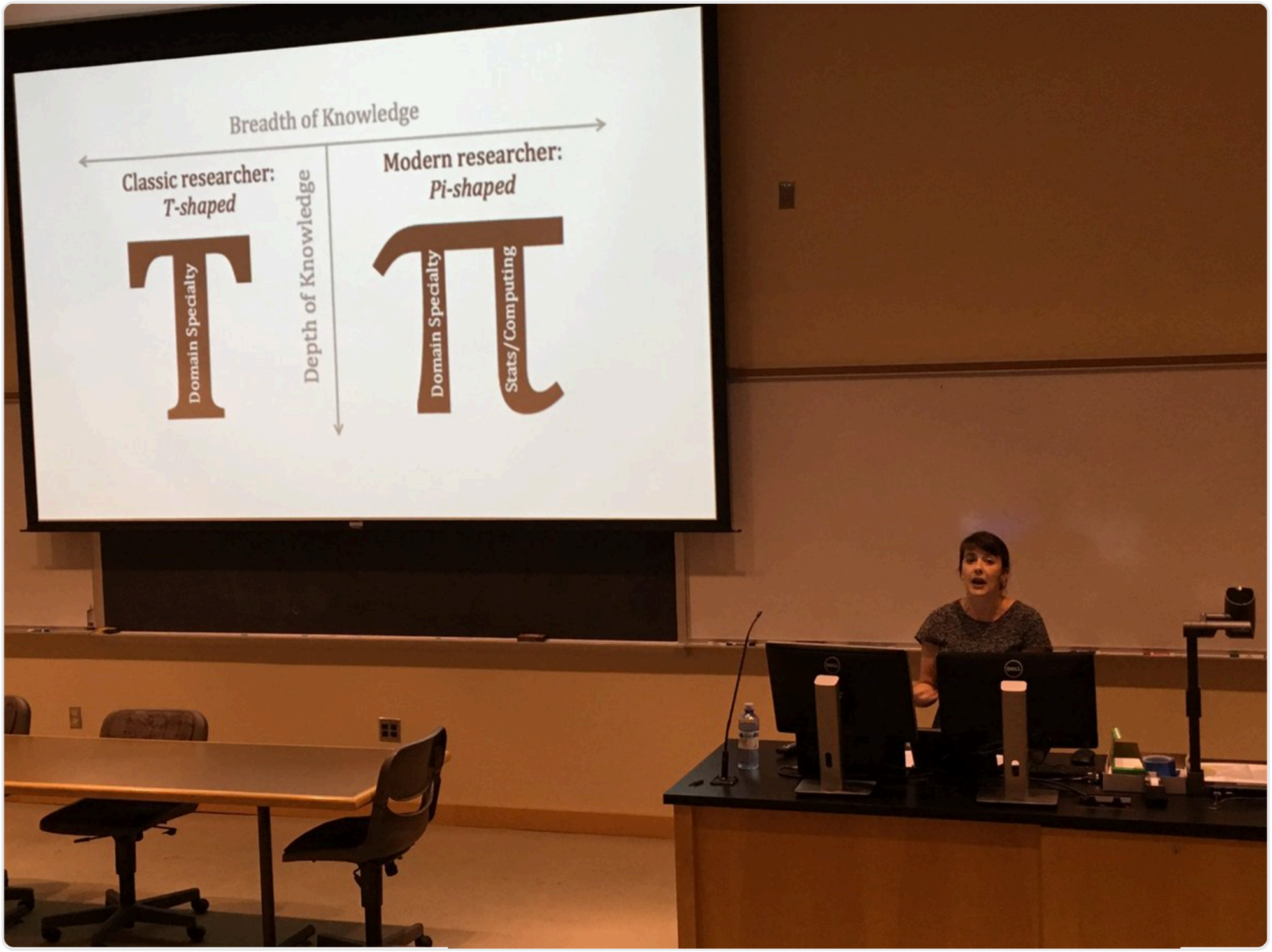
Base: Higher education: 127; Business: 63
Source: Gallup and BHEF, *Data Science and Analytics Higher Education Survey* (December 2016).

source pwc via @mikequindazzi

How can data science help professionals ?

Nicholas Horton @askdrstats · Sep 5

#ethnography of #datascience talk by Brittany Fiore-Garland challenges and opportunities @BrittaFiore @LafCol @ThisisStats #statistics



Early/Mid Career

- Decision making assistance
  - All companies collect data
  - Many struggle with representing data
  - Presentation of status updates
  - A vs B , why ?
  - Applicable across domains

On the job activities ..

PLOTS

REPORTING

ANIMATIONS

DATA VISUALIZATION

REPRODUCIBLE REPORTS

DATA COLLECTION

DATA CLEANING

DATA IMPORTING

WRITING TEMPLATES

Higher studies ?

- Research
  - Writing journal papers
  - Conference presentations
- Part time jobs
  - Data science skills are in high demand
  - Research assistantship
  - Teaching assistantship
  - Grant writing

As you enter C-Suite

- Dashboards
- Decision Science
- Cross validation
- Board meeting presentations
- Corporate Social Responsibility Initiatives

STORY TELLING

Elements of a Data Story

Problem

Data ->
Visualizations ->
Emotions ->
Decisions ->
Actions

Solution

Grammar of graphics

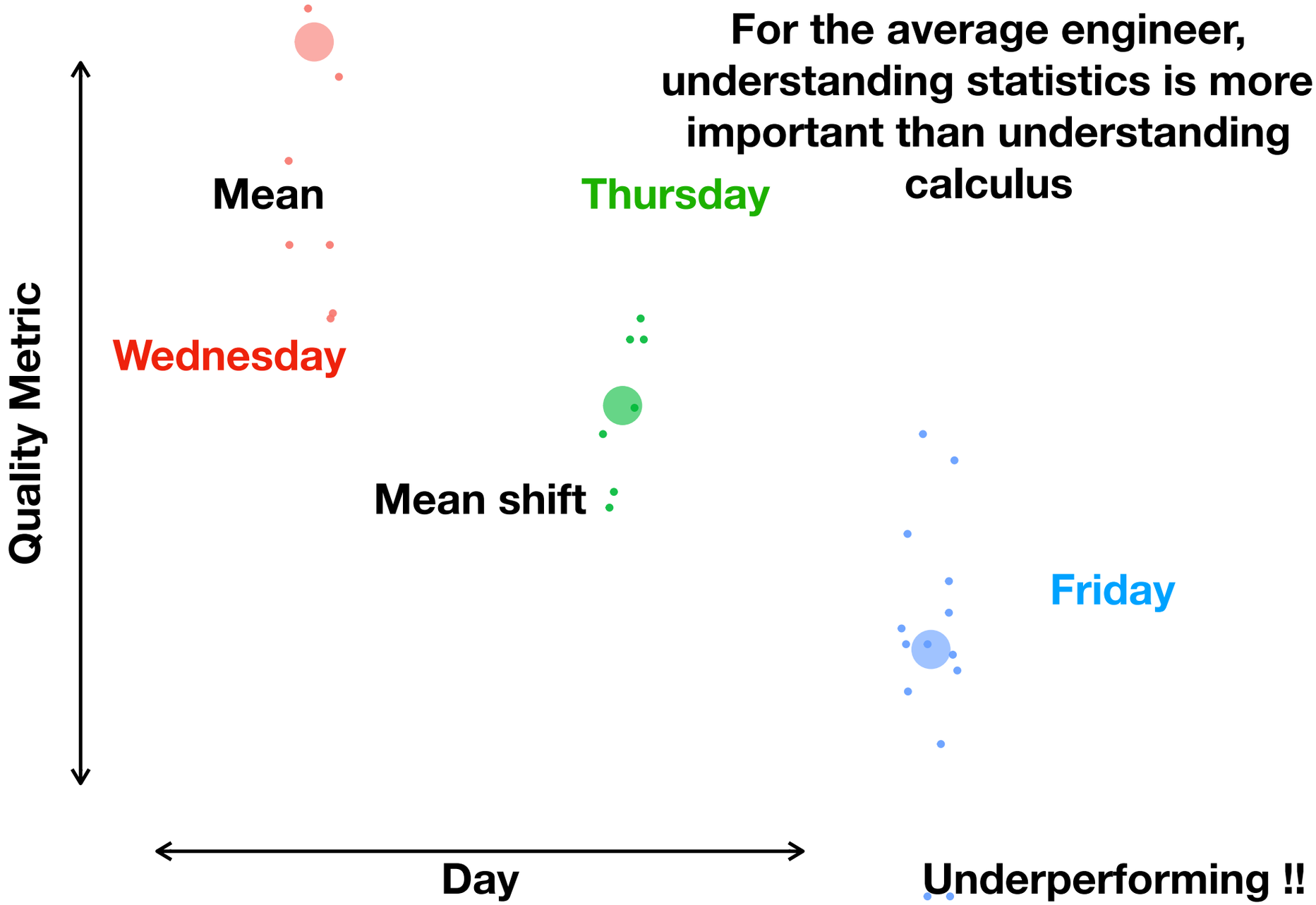
- Aesthetic mappings
  - x, y, colour, size, shape etc
- Geoms
  - Points, lines, polygons etc

Categorical comparisons

Outlier : good or bad ?

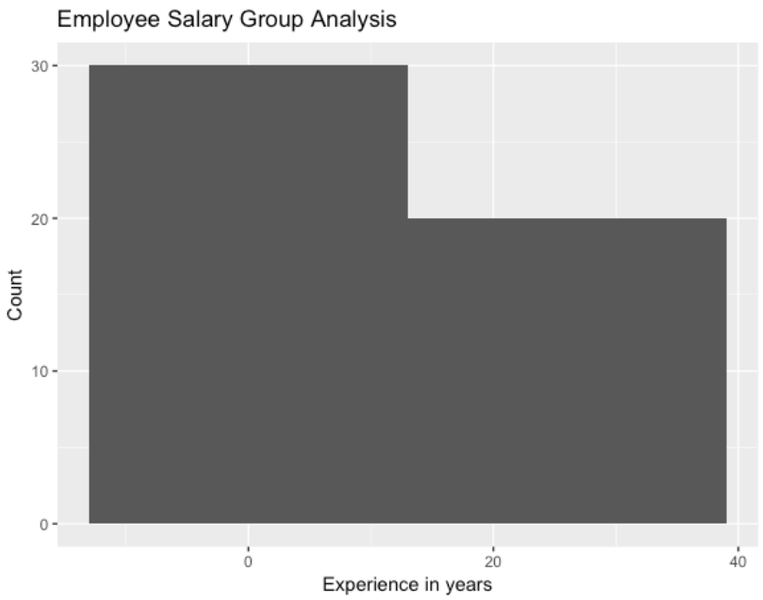
Moment of truth :

For the average engineer, understanding statistics is more important than understanding calculus

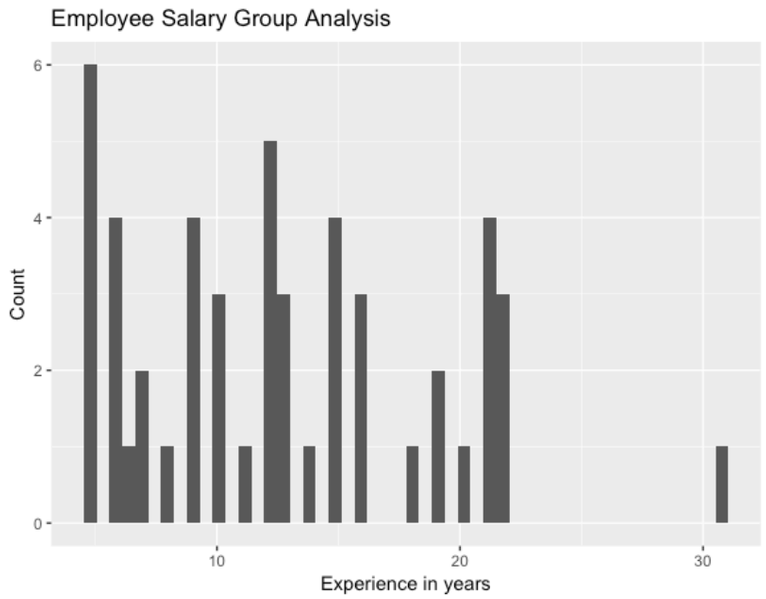


1-variable(numeric) histogram

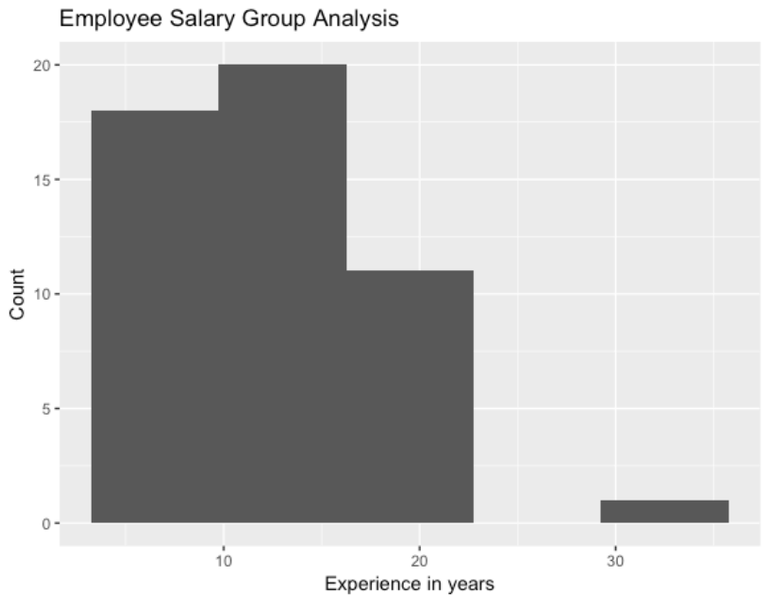
Histograms



Low bin count :
Broad grouping of salaries
How many high paid vs low paid employees
Less useful



High bin count :
Individual differences highlighted
Out-liers will stand out
Data range can be inferred

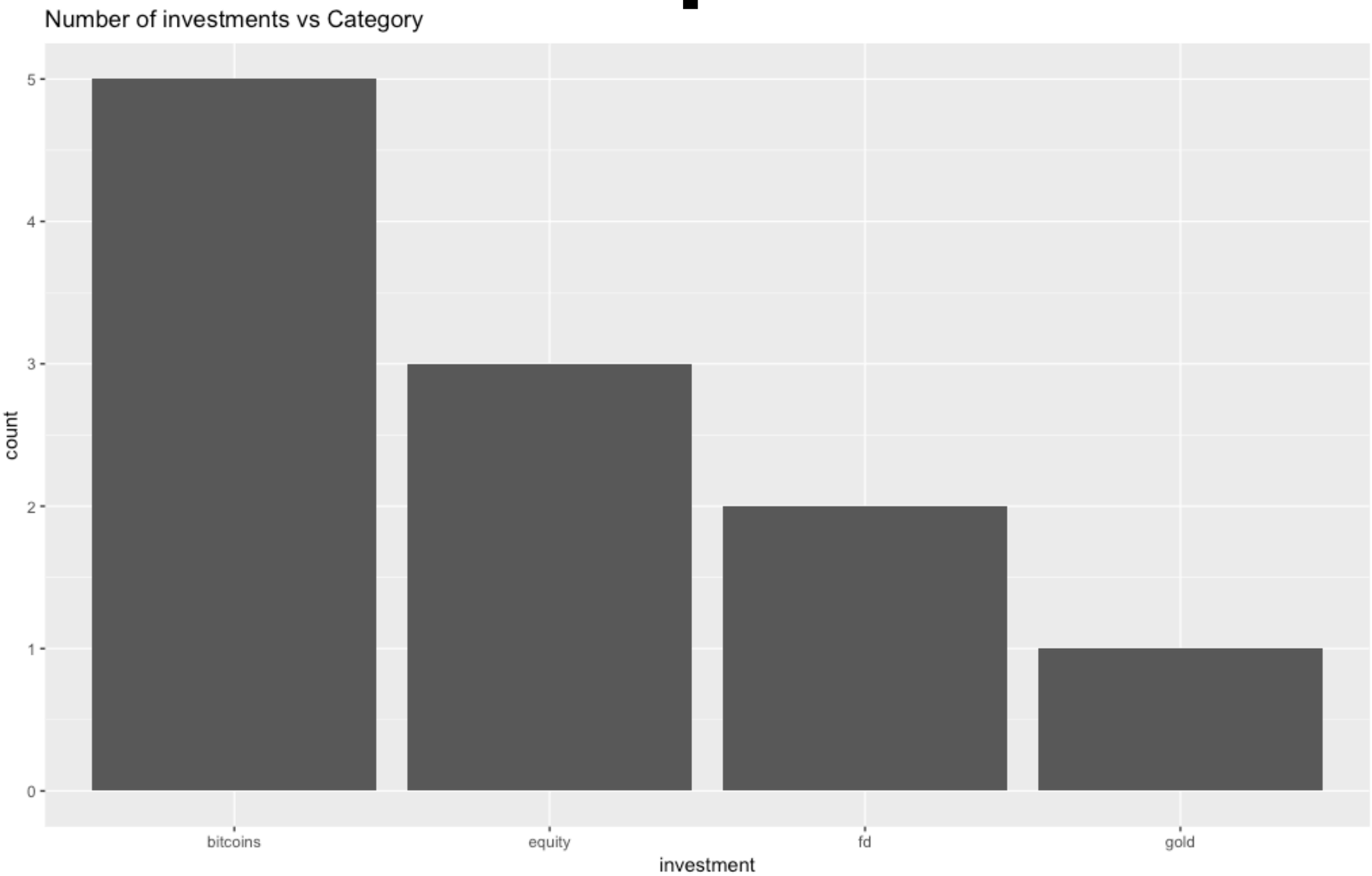


Medium bin count :
Identify salary groups in the data if any
Group comparisons highlighted
Can identify where data peaks
Outliers may or may not stand out

1-variable(categorical)

barplot

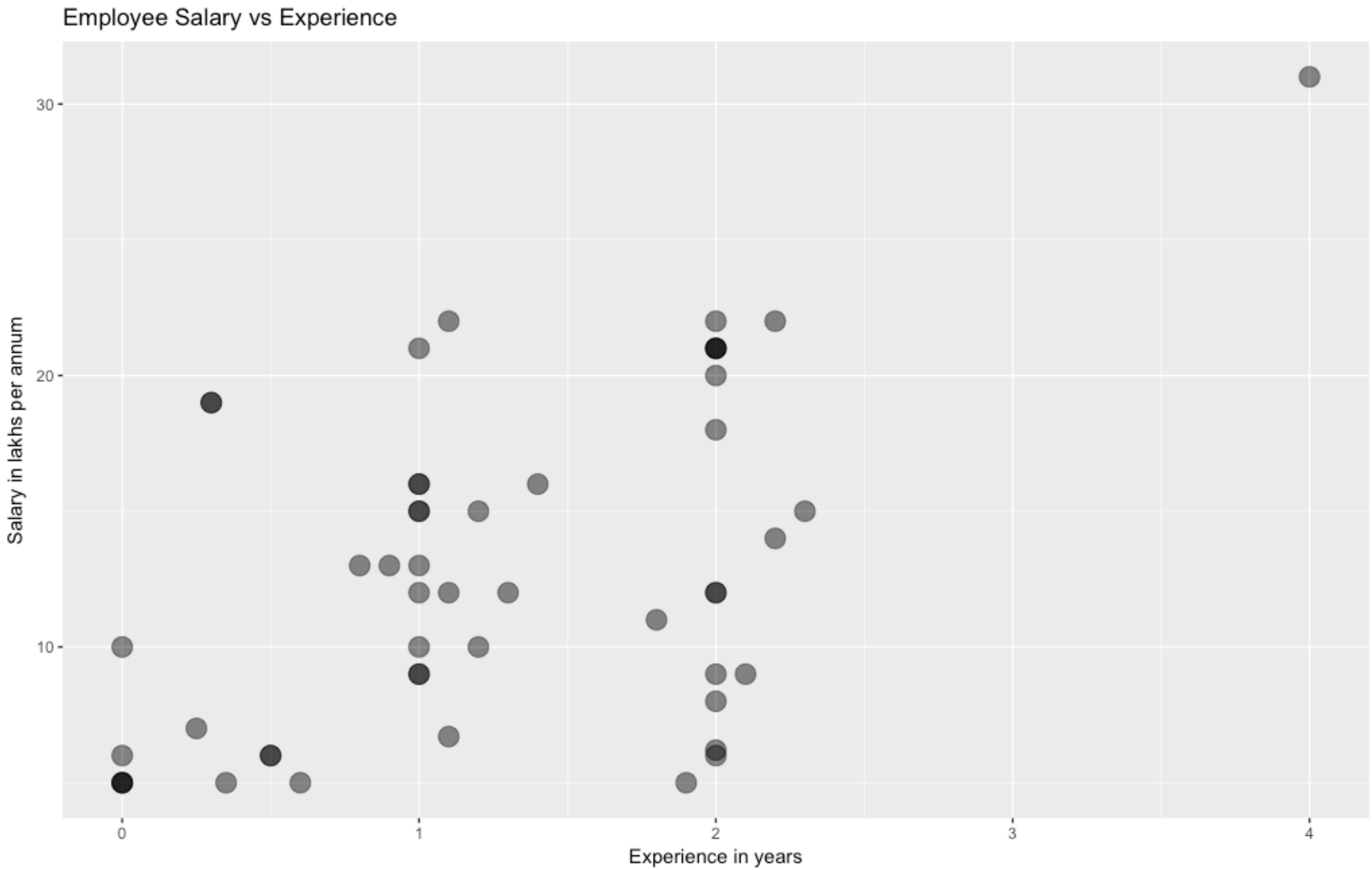
Barplot


Number of investments vs Category
count
5
4
3
2
1
0
bitcoins
equity
fd
gold
investment

2-variables (num vs num)

Scatterplot

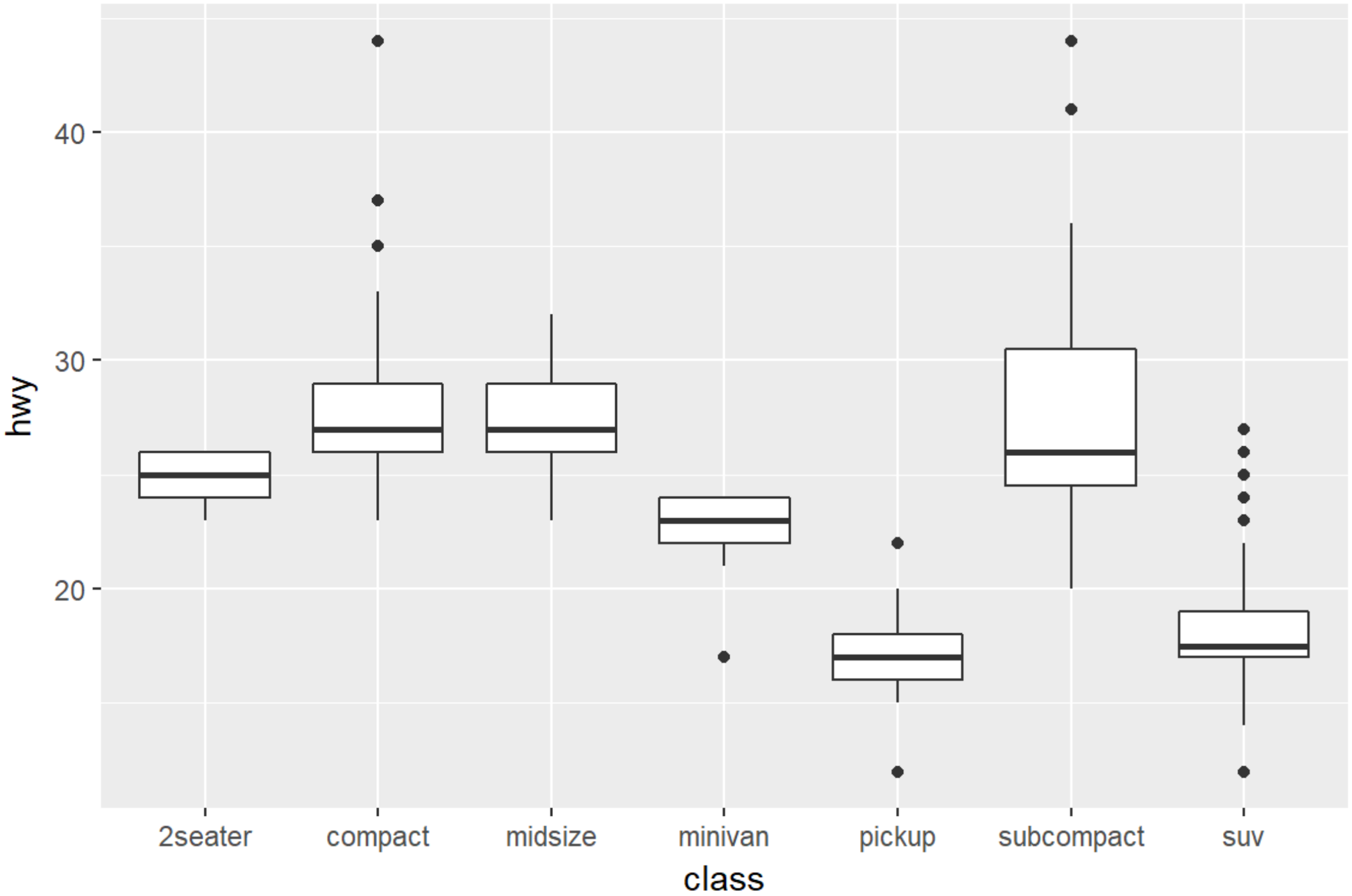
Scatterplot



2-variables (num vs categorical)

boxplot

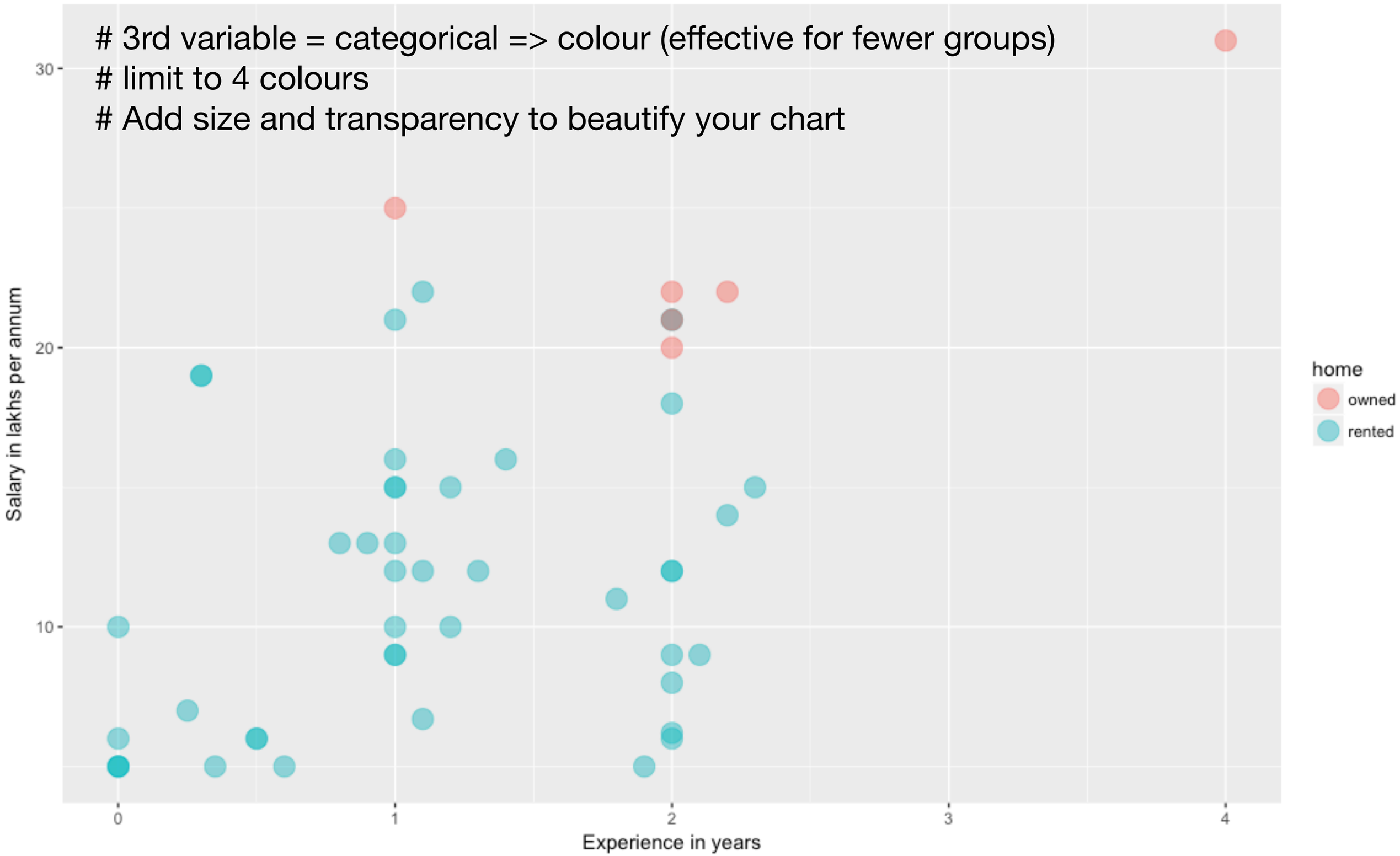
Boxplot


40
30
20
hwy
2seater
compact
midsize
minivan
pickup
subcompact
suv
class

3-variables

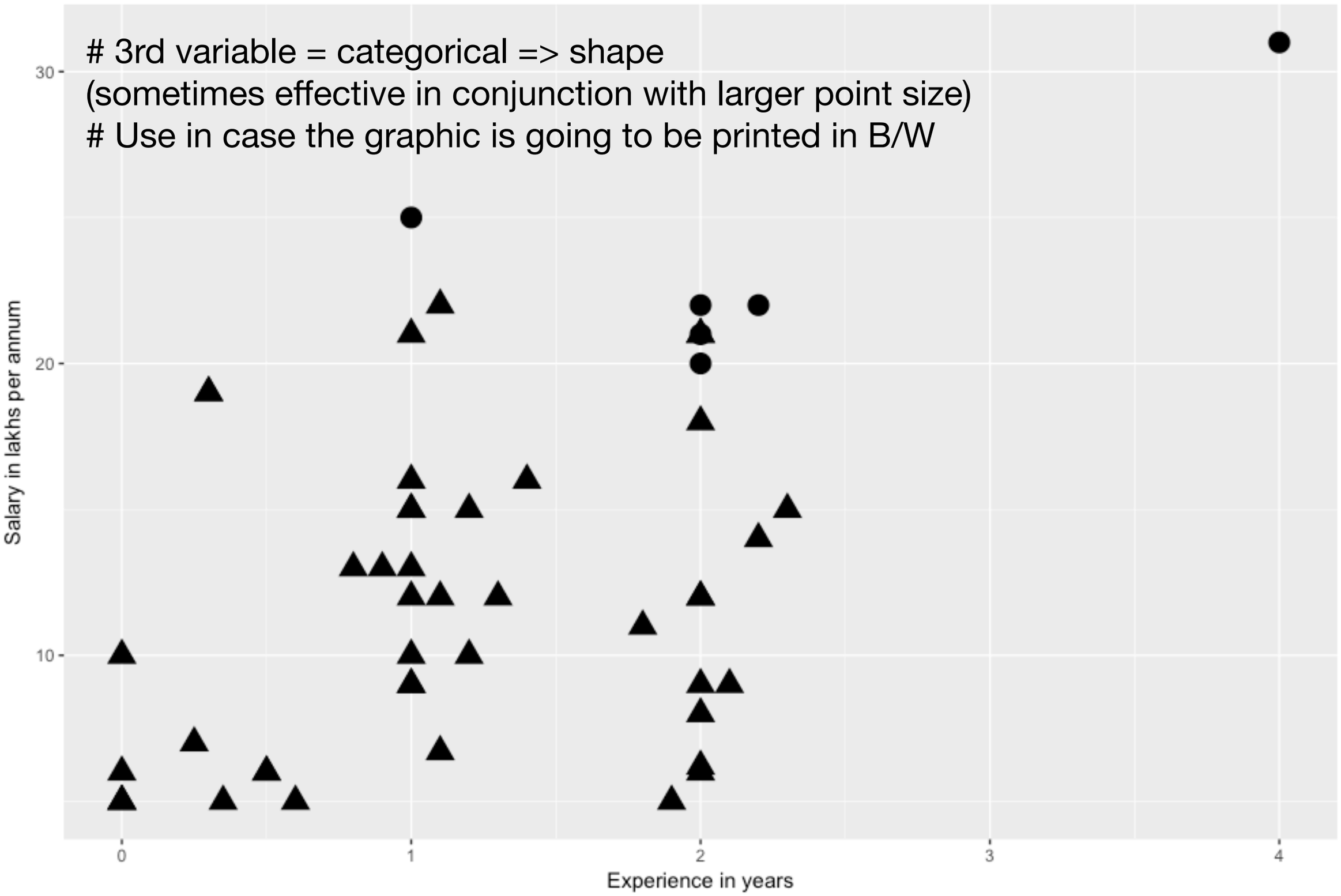
Employee Salary vs Experience vs Home Ownership

3rd variable = categorical => colour (effective for fewer groups)
limit to 4 colours
Add size and transparency to beautify your chart

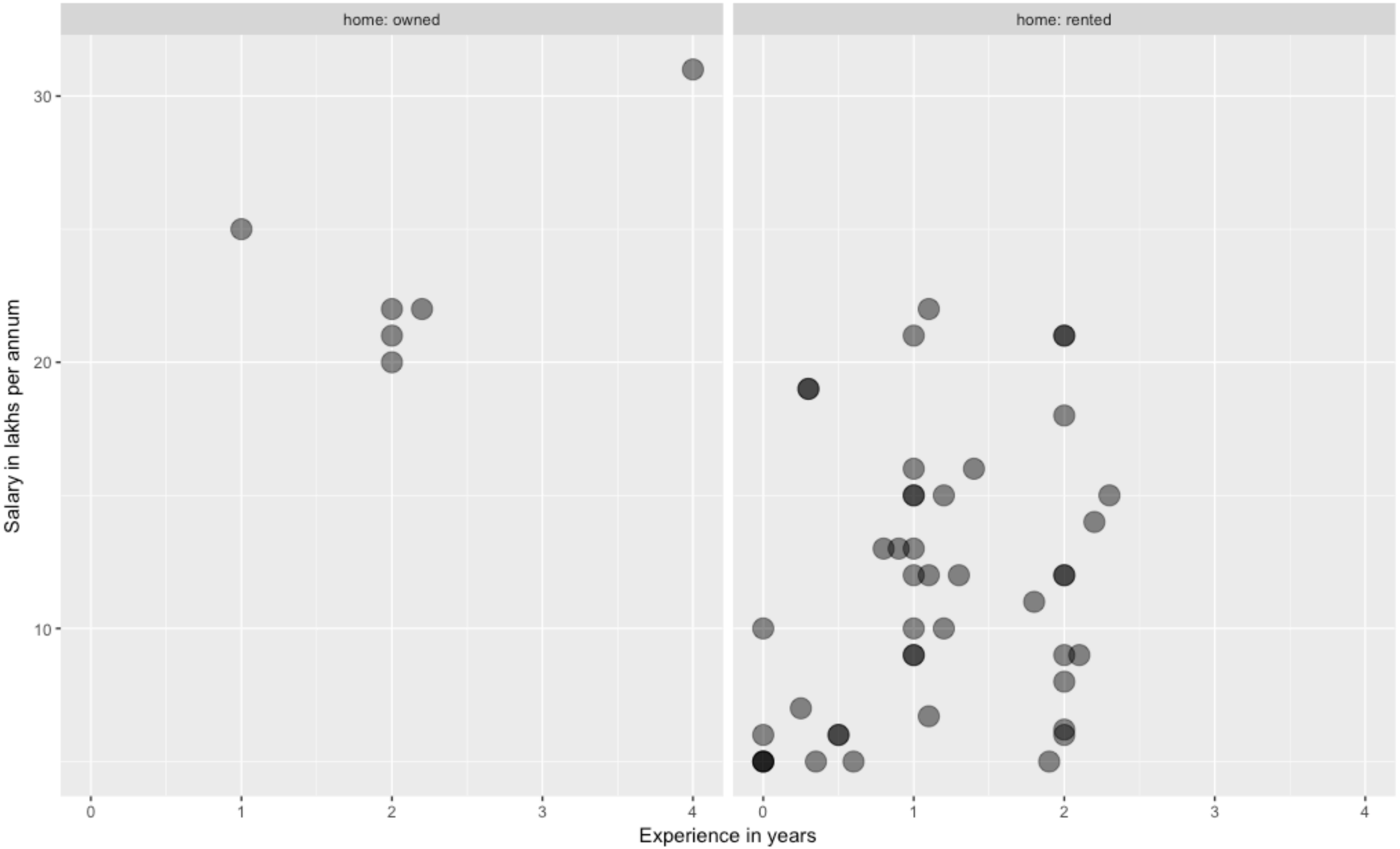


Employee Salary vs Experience vs Home Ownership

3rd variable = categorical => shape
(sometimes effective in conjunction with larger point size)
Use in case the graphic is going to be printed in B/W



Employee Salary vs Experience vs Home Ownership



9-variable plot

Employee diversity data of a sporting app development company

Id	Name	Gender	Experience	Position	Location	Salary	Sport	Travel
SR01	Sachin	M	2	Engineer	India	10	Cricket	Will travel
SR02	Ram	M	4	Engineer	India	11	Cricket	Will travel
SR03	Jack	M	6	Manager	India	12	Baseball	Will travel
SR04	Tony	M	7	Consultant	USA	15	Baseball	Will travel
SR05	Pavani	F	1	Engineer	Switzerland	5	Hockey	Will travel
SR06	Leela	F	9	Manager	India	20	Hockey	Will travel
SR07	Roger	M	2	Engineer	Switzerland	36	Soccer	Will not travel
SR08	Catherine	F	1	Engineer	Switzerland	38	Soccer	Will not travel
SR09	Ismail	M	5	Manager	India	10	Soccer	Will travel
SR10	Carla	F	20	CEO	India	50	Soccer	Will travel
SR11	Alison	F	8	Manager	USA	40	Soccer	Will travel
SR12	Alia	F	5	Manager	India	13	Hockey	Will travel
SR13	Rekha	F	4	Consultant	India	14	Cricket	Will travel

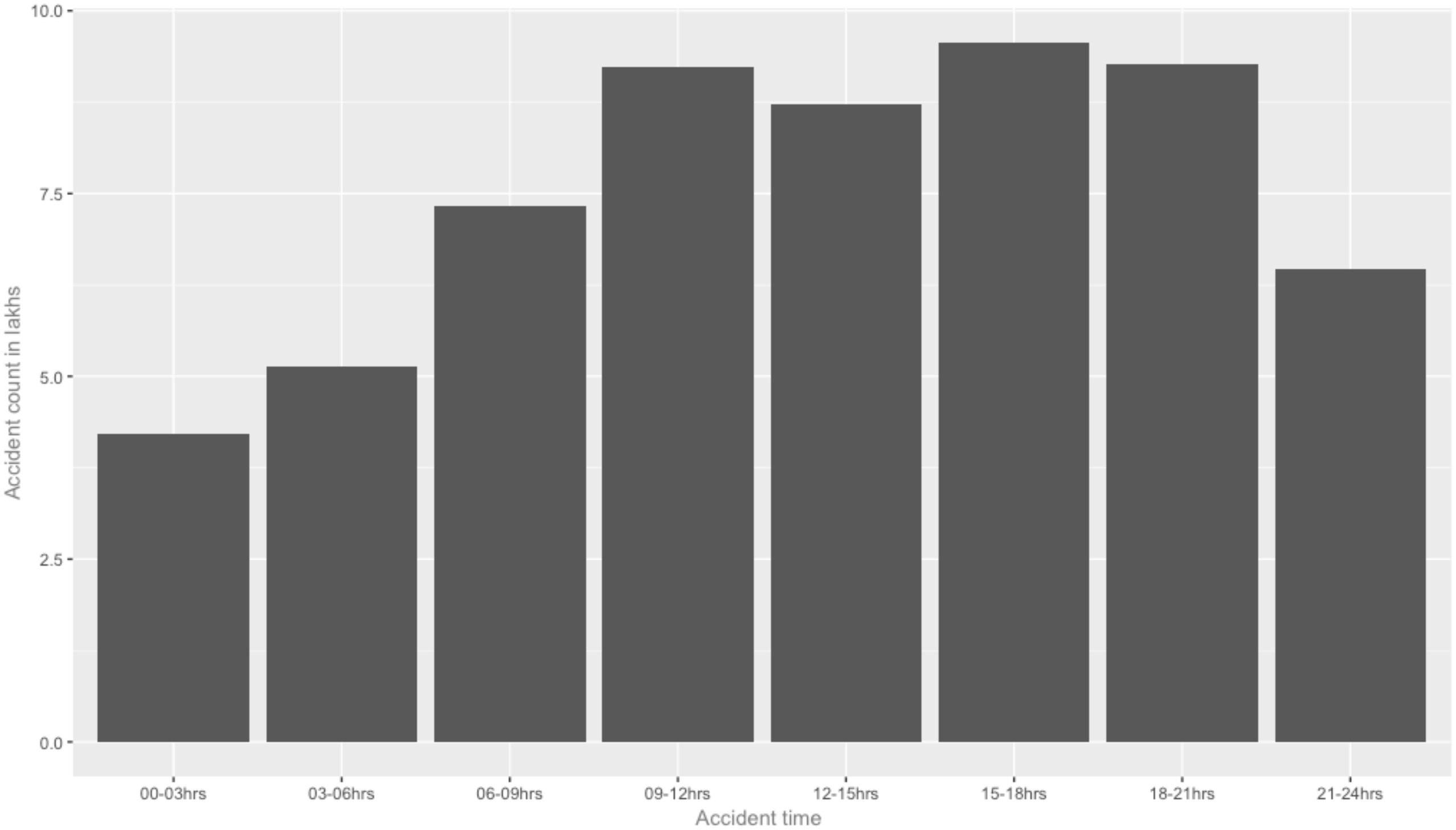
Employee diversity data

Case Study

Analyze accident trends in India

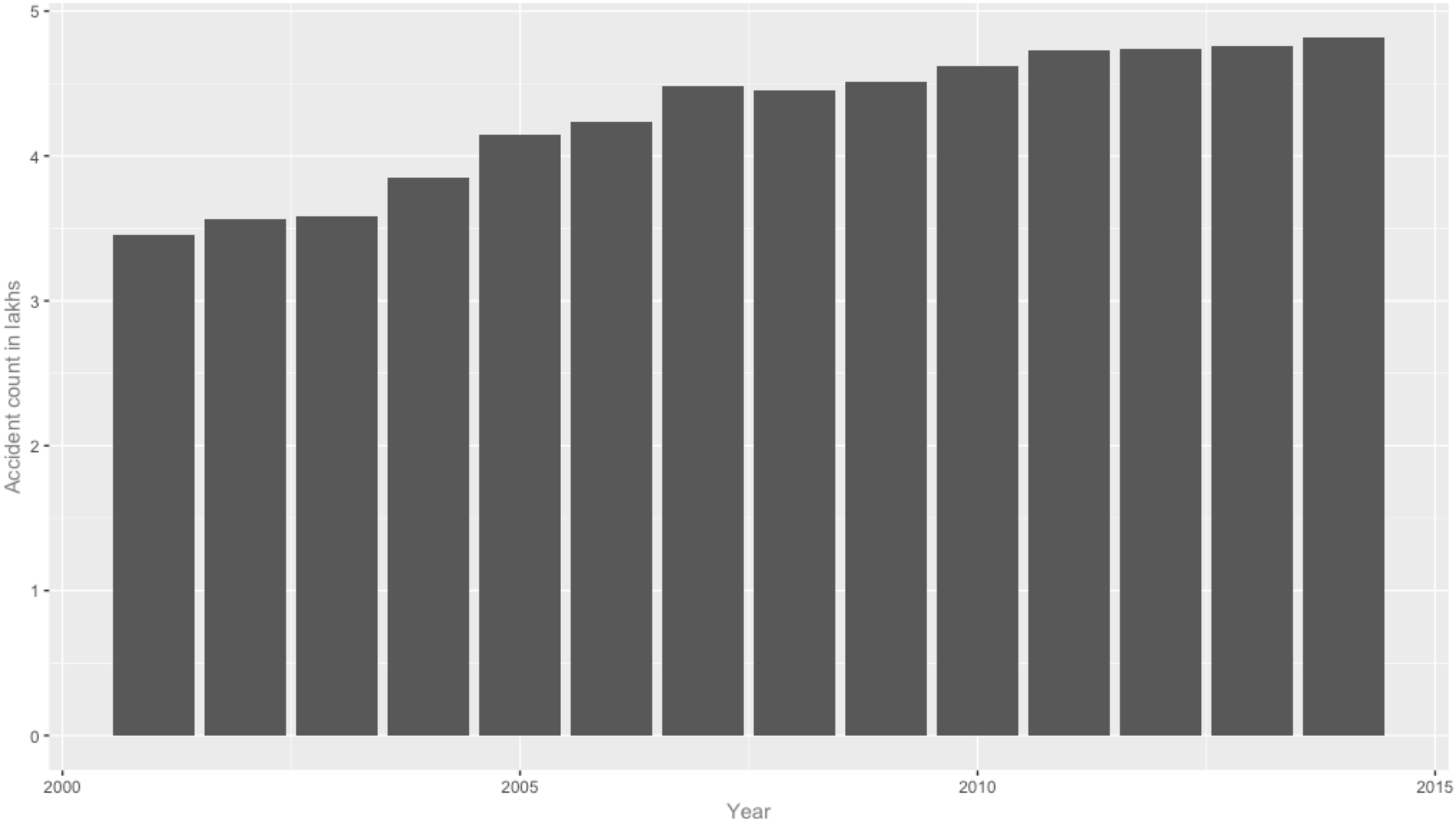
More accidents during day or night ?

Accidents in India (Time Interval)



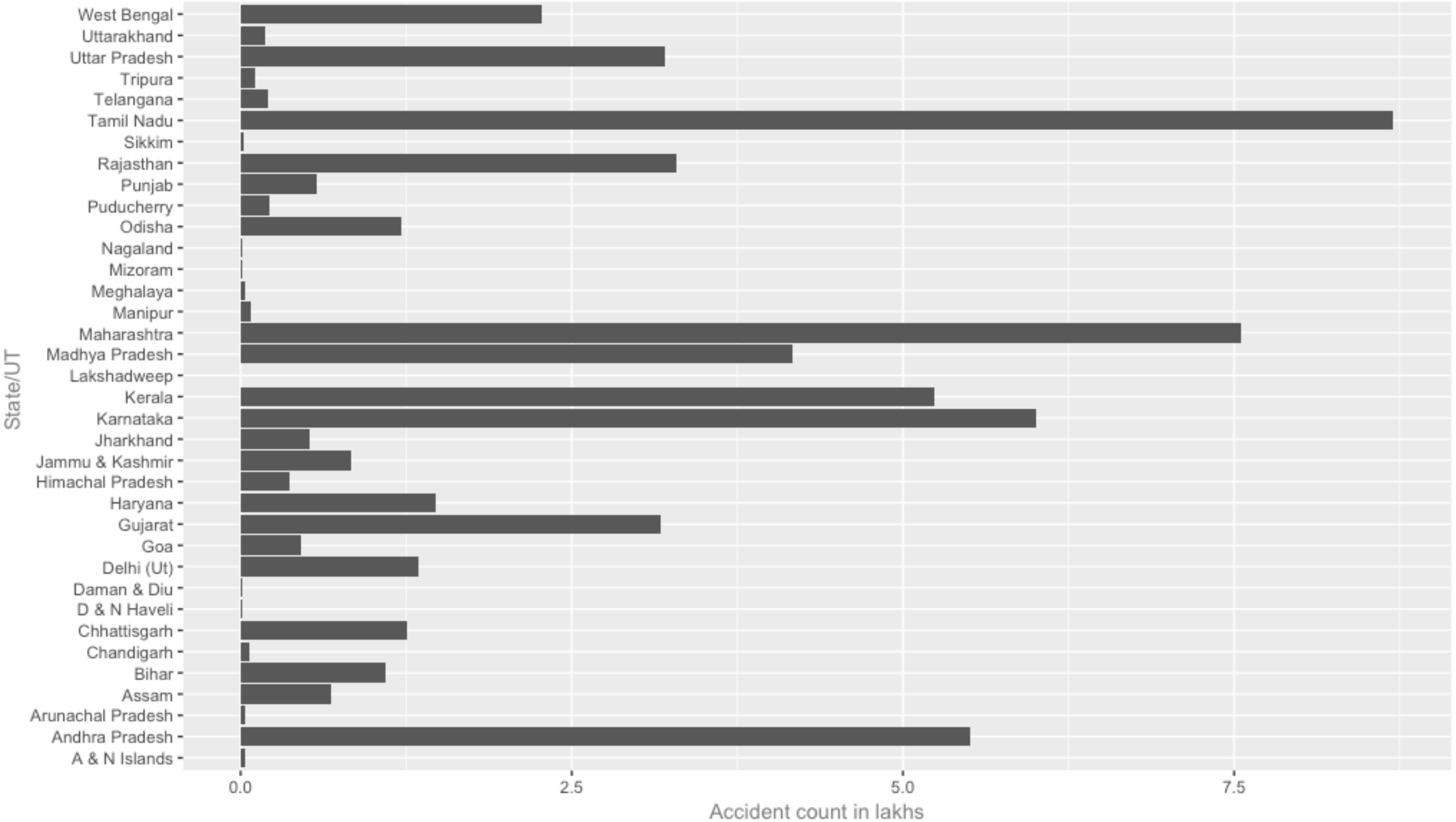
source : kaggle 2001-2014

Accidents in India (Yearly Trend)

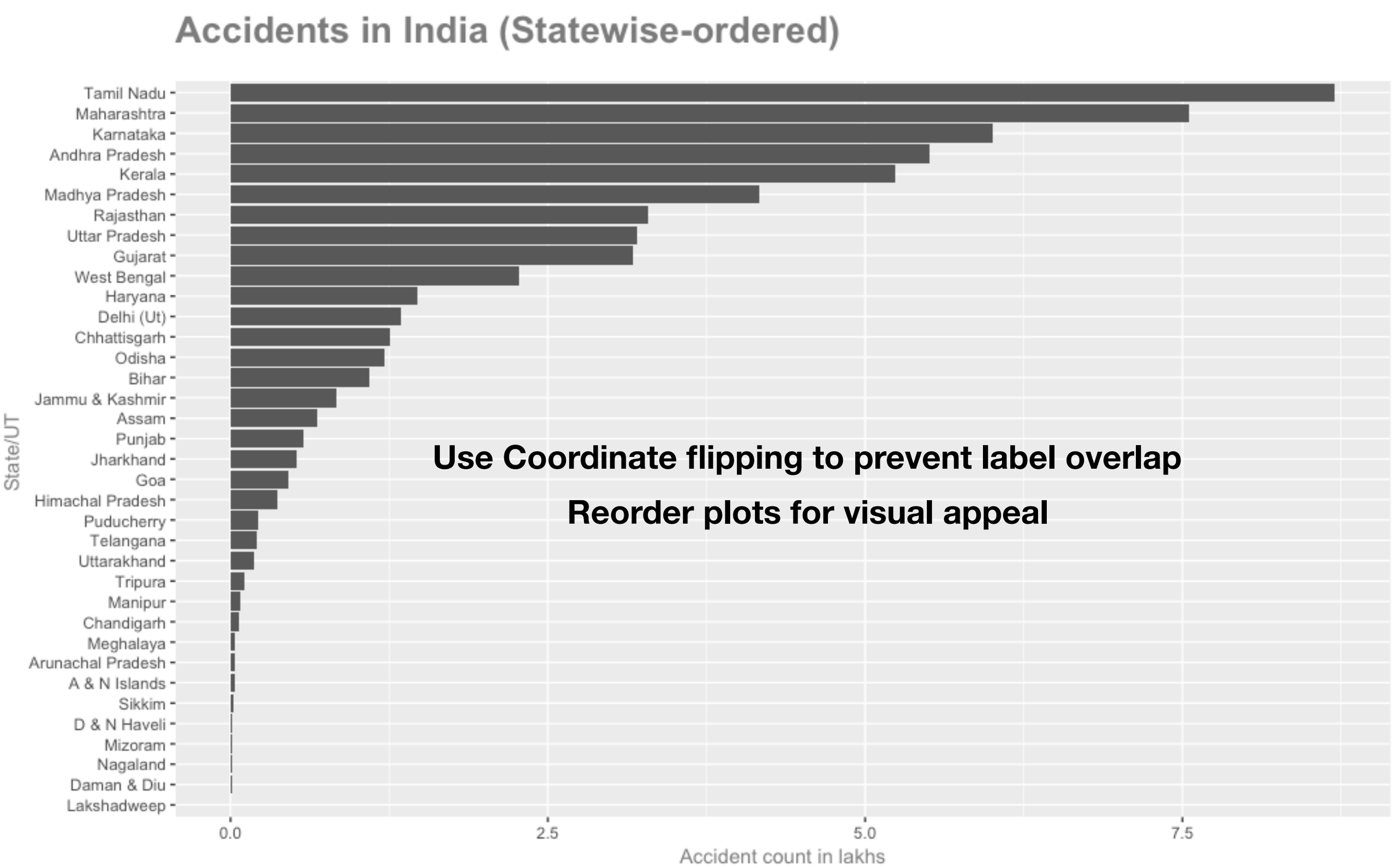


source : kaggle 2001-2014

Accidents in India (Statewise)



source : kaggle 2001-2014


Accidents in India (Statewise-ordered)
State/UT
Tamil Nadu
Maharashtra
Karnataka
Andhra Pradesh
Kerala
Madhya Pradesh
Rajasthan
Uttar Pradesh
Gujarat
West Bengal
Haryana
Delhi (Ut)
Chhattisgarh
Odisha
Bihar
Jammu & Kashmir
Assam
Punjab
Jharkhand
Goa
Himachal Pradesh
Puducherry
Telangana
Uttarakhand
Tripura
Manipur
Chandigarh
Meghalaya
Arunachal Pradesh
A & N Islands
Sikkim
D & N Haveli
Mizoram
Nagaland
Daman & Diu
Lakshadweep
0.0
2.5
5.0
7.5
Accident count in lakhs
Use Coordinate flipping to prevent label overlap
Reorder plots for visual appeal
source : kaggle 2001-2014

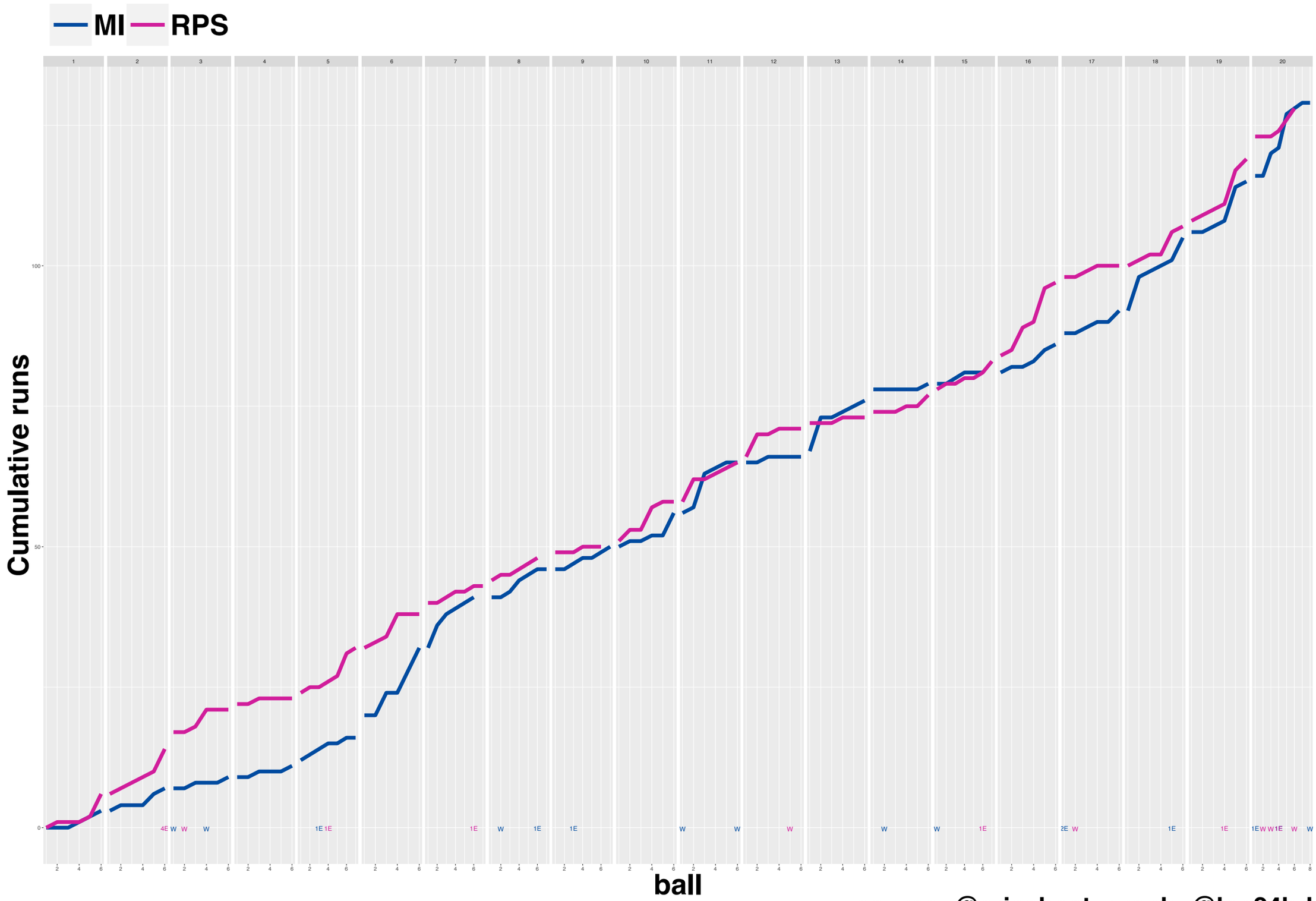
Case Study

ipl_final_viz.R

Information Visualization !

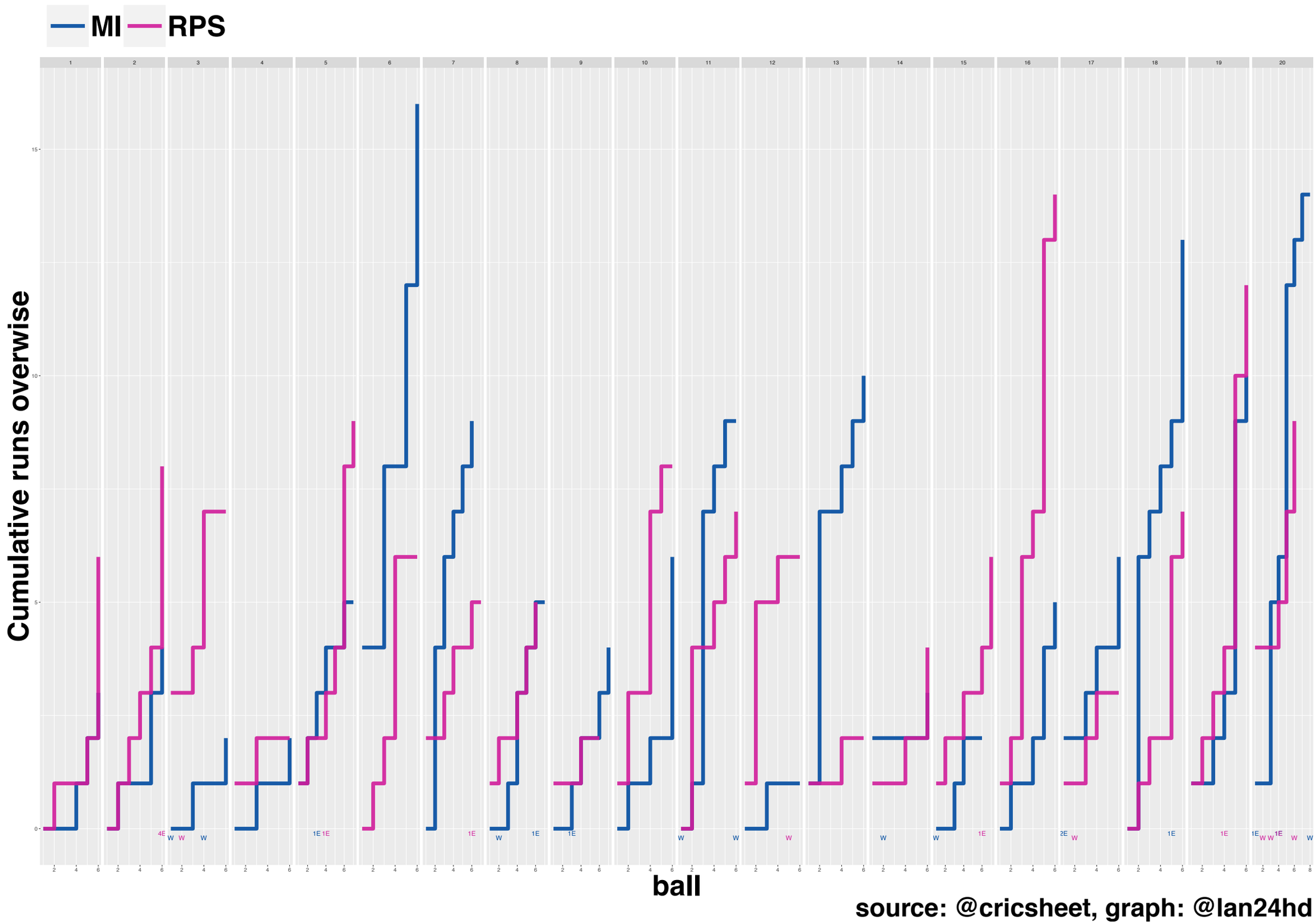
How did RPS lose to MI by just 1 run at the 2017 IPL T20 final ?

IPL2017 T20 Final – Runs Distribution

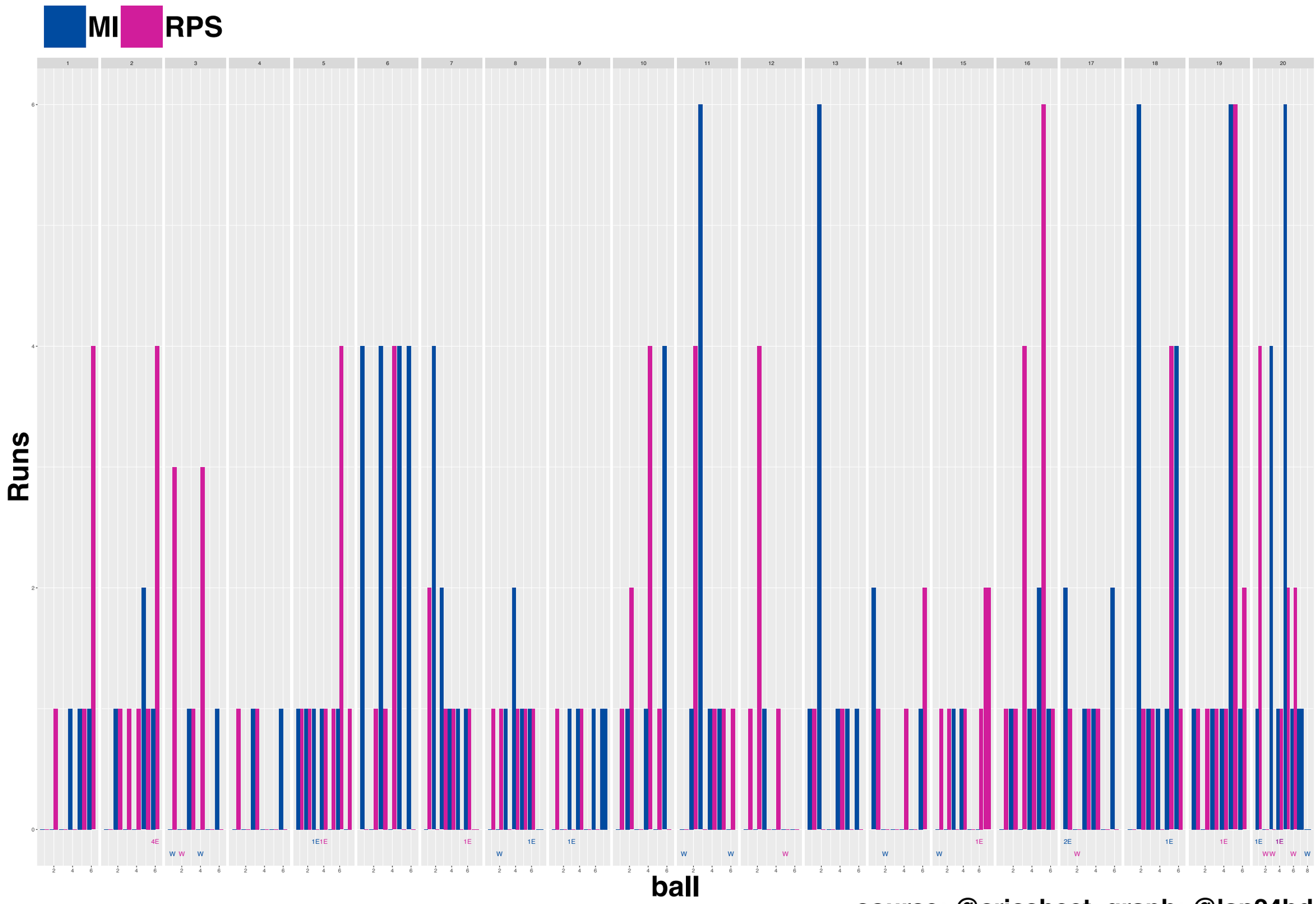


source: @cricsheet, graph: @Ian24hd

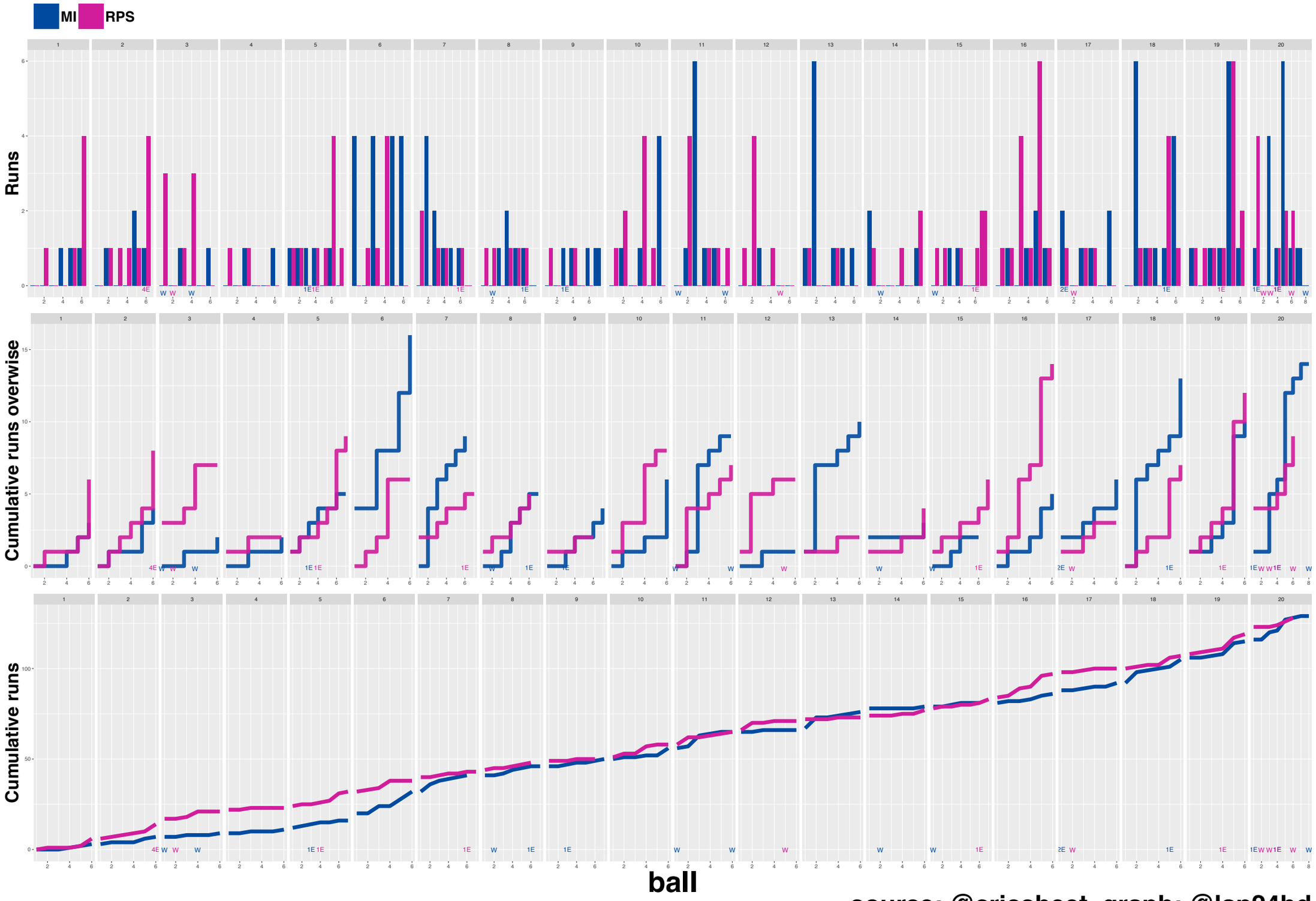
IPL2017 T20 Final – Runs Distribution



IPL2017 T20 Final – Runs Distribution



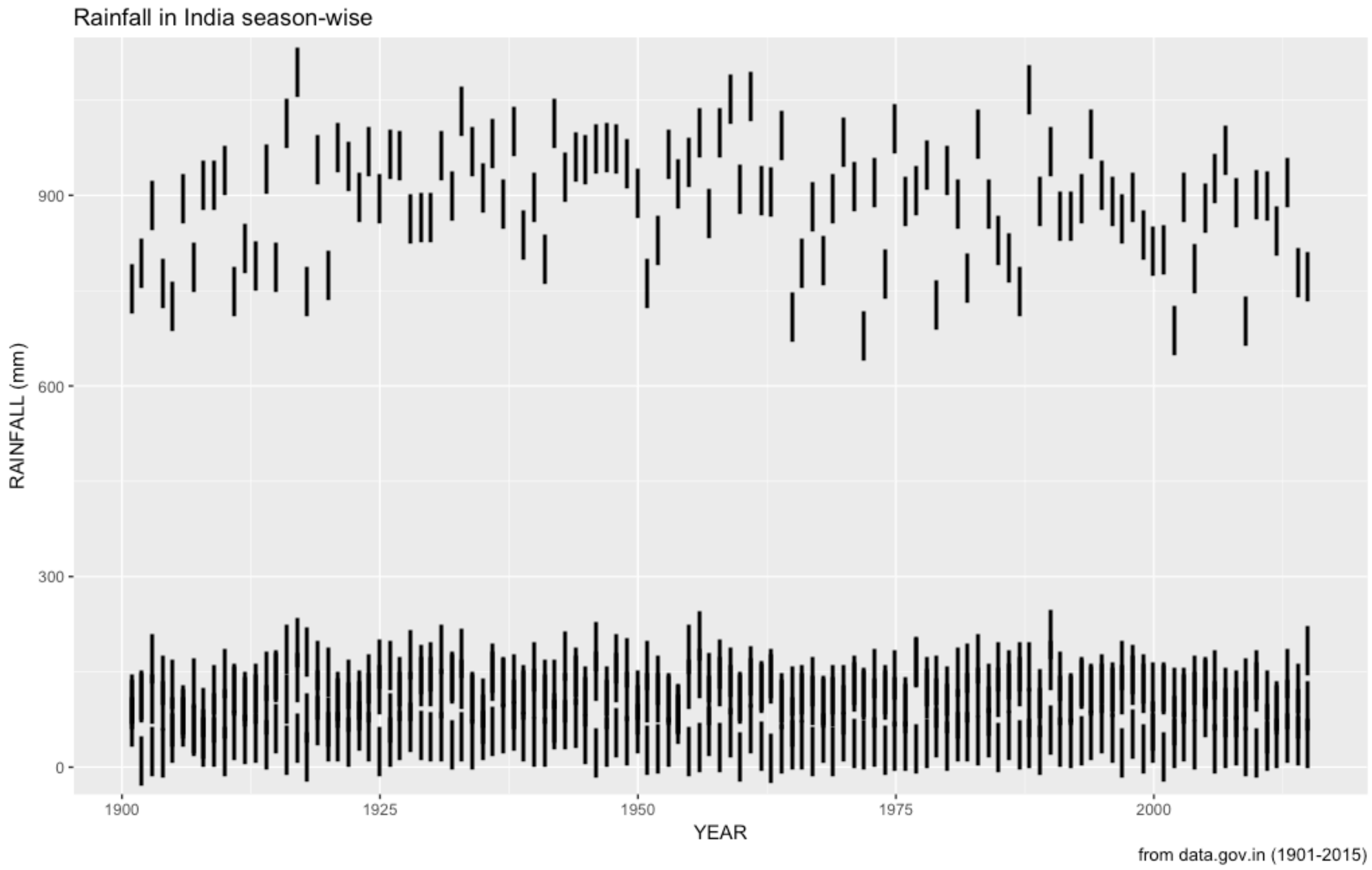
IPL2017 T20 Final – Runs Distribution

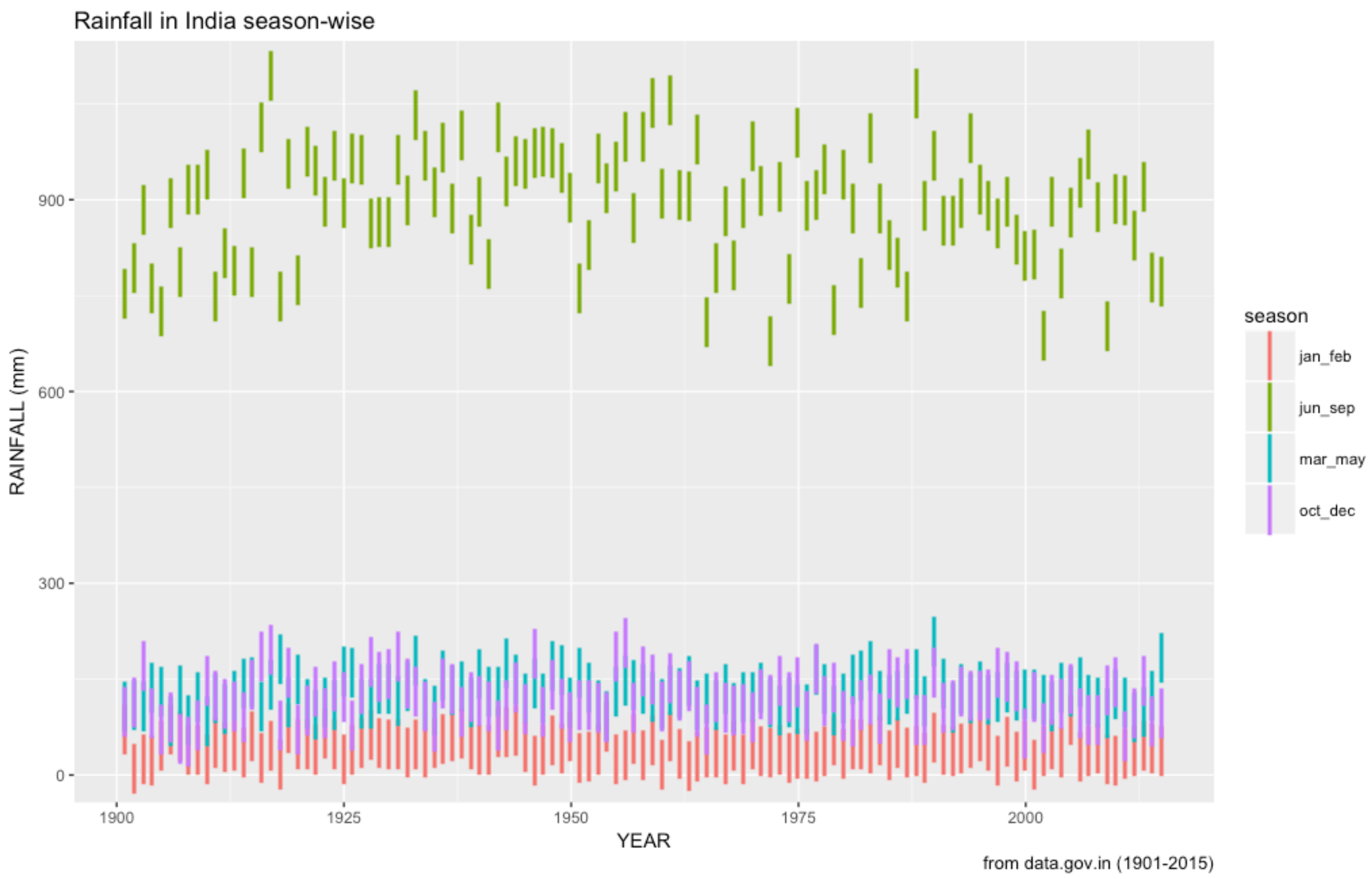


source: @cricsheet, graph: @Ian24hd

Case Study

Visualize Rainfall in India 1901-2015

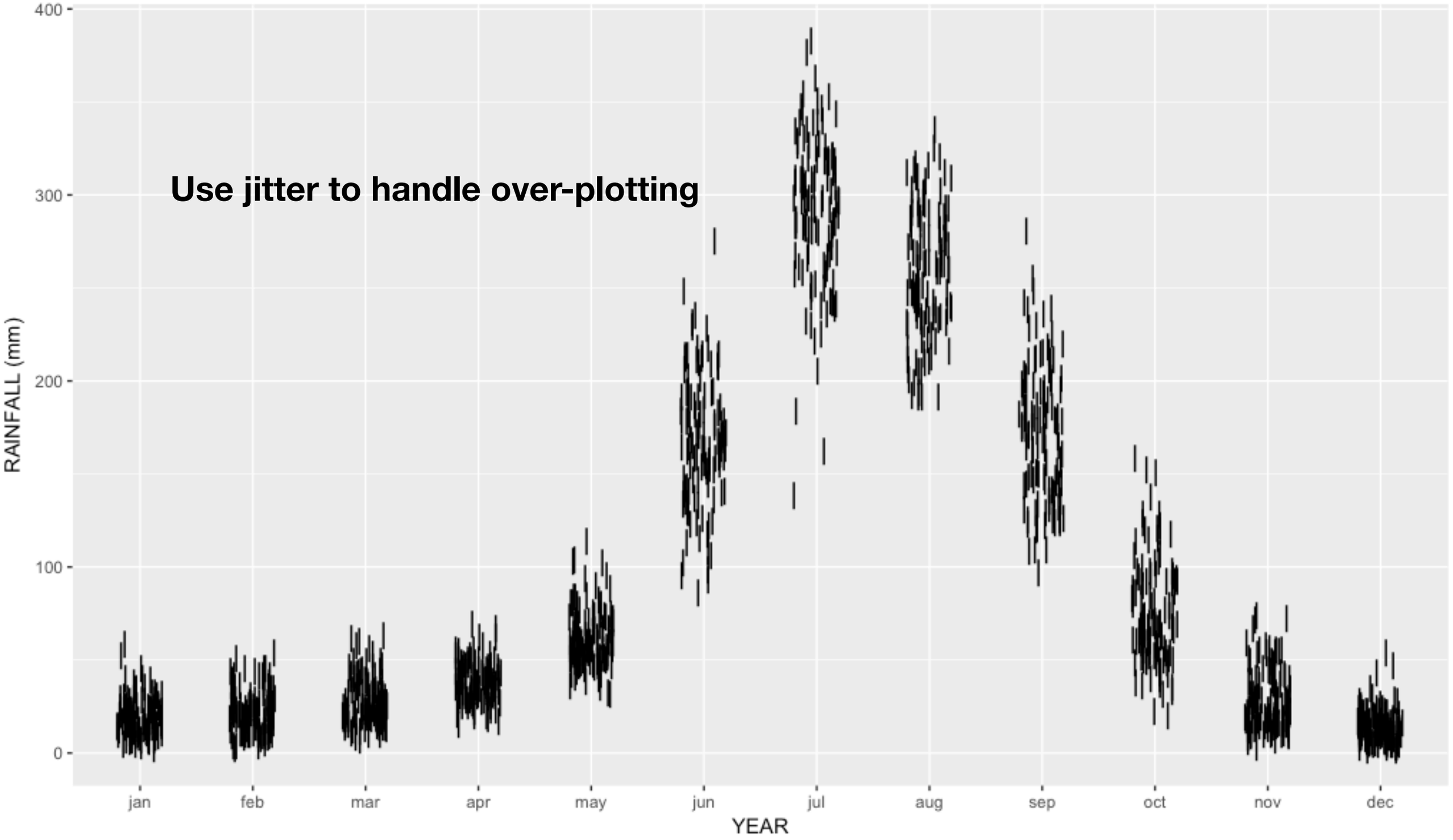

Rainfall in India season-wise
RAINFALL (mm)
900
600
300
0
1900
1925
1950
1975
2000
YEAR
from data.gov.in (1901-2015)

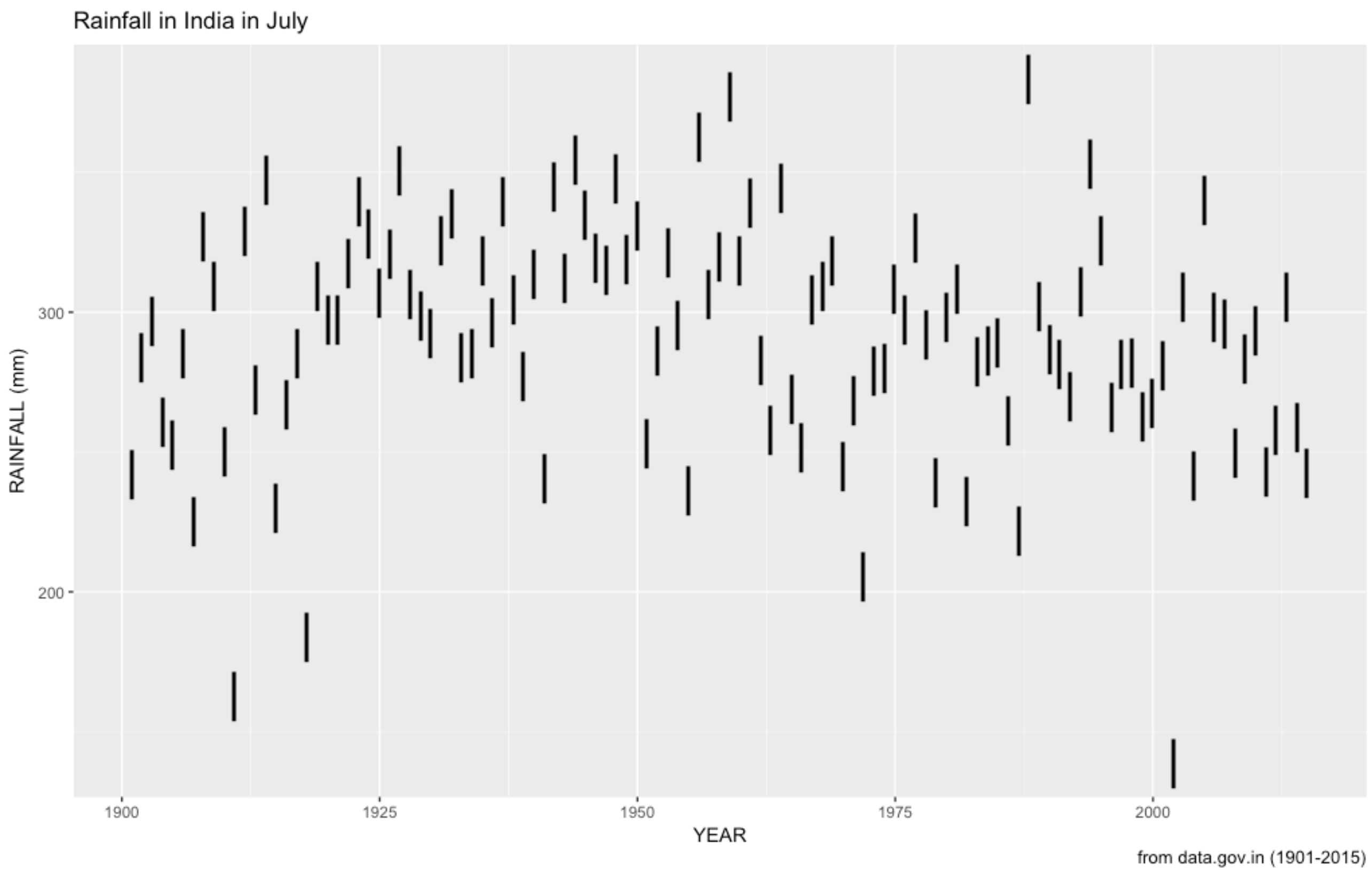

Rainfall in India season-wise
RAINFALL (mm)
900
600
300
0
1900
1925
1950
1975
2000
YEAR
from data.gov.in (1901-2015)
season
jan_feb
jun_sep
mar_may
oct_dec

Which is the wettest month in India ?

Rainfall in India

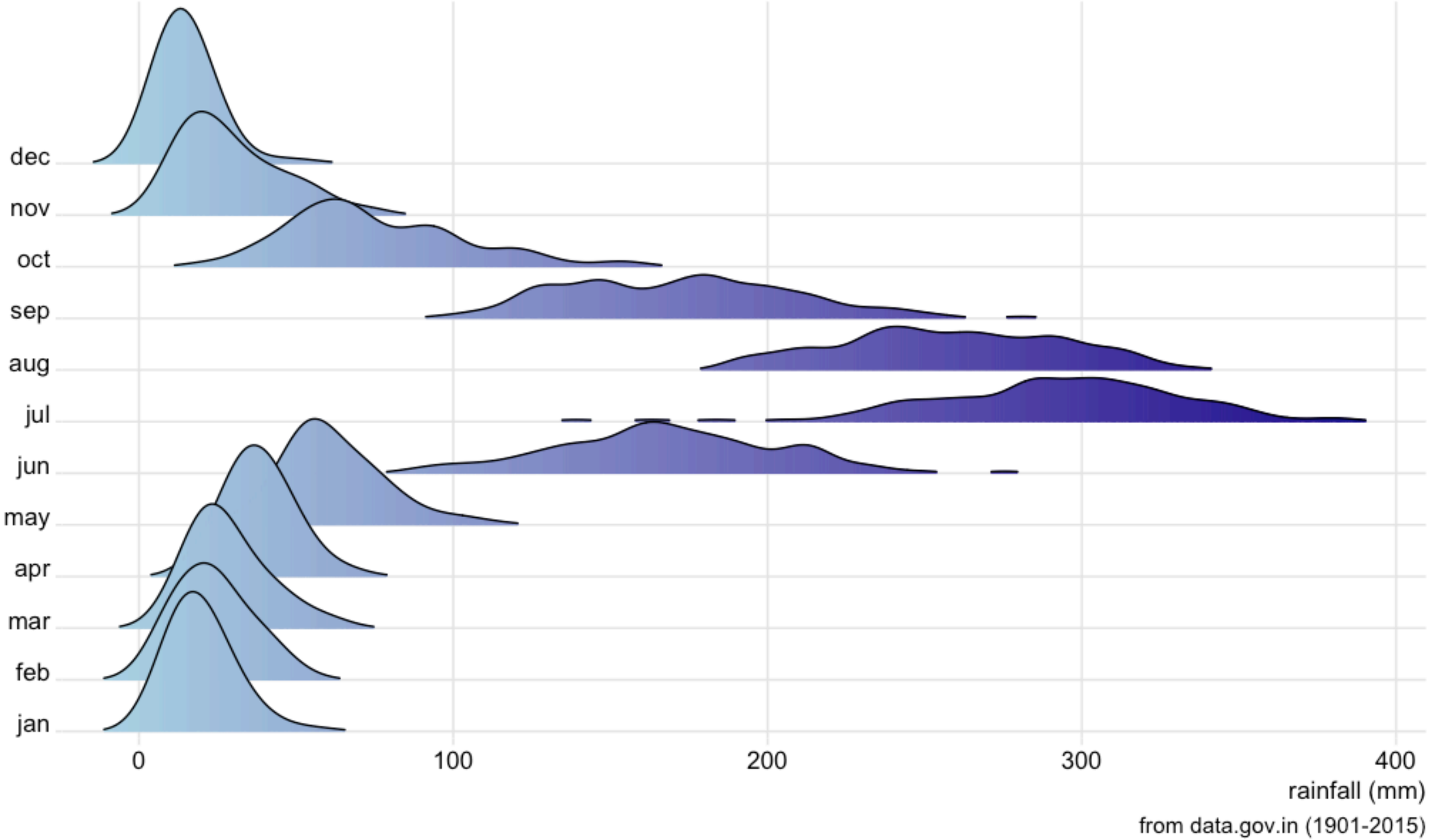
Use jitter to handle over-plotting




Rainfall in India in July
RAINFALL (mm)
300
200
1900
1925
1950
1975
2000
YEAR
from data.gov.in (1901-2015)

Rainfall in India month-wise distribution (joyplot/ridges plot)

One of the latest developments in the field of data viz : joyplot/ridges plot for numeric distributions !



Summary Tips

Before plotting !

- Decide : Simple plots for decision makers vs Complex plots for analysts
- Start with raw data and clean up names
- Data reshaping (wide to long)
- Handle missing values
- Data wrangling

When to choose what kind of plot ?

1 variable	Numeric	Histogram
1 variable	Categorical	Barplot
2 variables	Numeric vs Numeric	Scatterplot
2 variables	Numerical vs Categorical	Boxplot
3 variables	Color or shape or facet	
> 3 variables	Color and/or shape and/or facet	

Know your data !

- Know how to import data into your favourite programming language
- Multiple data sources ? How and when to merge them ?
- Get familiar with your data, its dimensions, data types : text or numeric or time series etc
- Raw vs Summary datasets ; Are they mixed? If so how to separate them
- Analyse subsets of data : Are there groups , are there patterns in your datasets
- How to add extra columns, arrange, and summarise data ?

Clean your data !

- What is tidy data ?
- Think variables and columns for each variables.
- Wide data (Human friendly) vs long data (computer friendly)
- Missing values : Drop or replace ?
- Spelling mistakes ?
- Case mismatches ?

Plot your data !

- “Visualizations can surprise you !” - Dr. Hadley Wickham
- Ah ! Now I see it. I did not expect that !
- What type of chart should I plot for the data at my hand : Histograms, Bar charts, Column charts, Scatter plots, Box plots ?
- Aesthetic mapping for your data : axes, colors, shapes, size and facet controls
- By-variable visual analysis !
- Data jittering, Coordinate flipping
- Arranging multiple plots in a predefined layout

Visual enhancement tips

- Use Colors + Shapes + Facets + Labels
- Minimize white space
- Reordering
- Data jittering
- Minimal theme
- Coordinate flipping

Let your data speak !

- Communicate your findings
- Reproducible reports (pdf, html)
- Interactive charts , web-enabled !
- Animations !

Data Visualization : Human Intelligence :: Machine Learning : Artificial Intelligence

Dr. Anand Lakshmanan
CEO, SIRPI PRODUCTS AND SERVICES PRIVATE LIMITED
@lan24hd
+91 83107 64903
anand@sirpi.io